DOCUMENT RESUME

ED 040 845 24 RE 003 050

AUTHOR	Harckham, Laura D.
TITLE	Development of Teacher Evaluation Scales to Predict Reading Success of Pupils in Primary Grades.
INSTITUTION	Fordham Univ., Bronx, N.Y.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-9-B-087
PUB DATE	Apr 70
GRANT	OEG-2-9-420087-1052(010)
NOTE	144p.
EDRS PRICE	EDRS Price MF-$0.75 HC-$7.30
DESCRIPTORS	Grade 1, Grade 2, Grade 3, Grade 4, Kindergarten Children, *Longitudinal Studies, *Predictive Measurement, *Reading Achievement, *Standardized Tests, Test Reliability

ABSTRACT

A longitudinal study was undertaken to determine whether measures that are readily available in most school districts, administered in kindergarten by classroom teachers, could predict reading achievement as measured by standardized tests in grades 1 through 4. The subjects were 553 children in the kindergarten class of 1964-65 in Ithaca, New York. They were given the Metropolitan Readiness Test, including the Goodenough Draw-A-Man Test; a ranking by the teacher; and a rating on a composite behavior rating scale. Achievement measures were the Metropolitan Achievement Test in grade 1 and the Stanford Achievement Test in grades 2 through 4. Among the findings were that the Metropolitan Readiness Test was the best single overall predictor, and Alphabet and Numbers subtests were the best subtest predictors. A bibliography and tables are included. (Author/NH)

ED040845

BR 9-B-087
PA 24
RE

FINAL REPORT
Project No. 9-B-087
Grant No. OEG-2-9-420087-1052(010)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

DEVELOPMENT OF TEACHER EVALUATION SCALES TO PREDICT READING SUCCESS OF PUPILS IN PRIMARY GRADES

Laura D. Harckham
Fordham University
Lincoln Center
New York, New York 10023

April 1970

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

RE003 050

Final Report
Project No. 9-B-087
Grant No. OEG-2-9-420087-1052(010)

DEVELOPMENT OF TEACHER EVALUATION SCALES TO PREDICT READING SUCCESS OF PUPILS IN PRIMARY GRADES

Laura D. Harckham

Fordham University

New York, New York

April 1970

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Statement of the Problem	4
	Definition of Terms	6
	Significance of the Problem	6
	Limitations of the Study	10
II.	REVIEW OF RELATED LITERATURE	11
	Readiness Tests	12
	Intelligence Measures	30
	Teacher Judgment	40
	Behavior Rating Scales	44
	Multivariate Studies	50
III.	THE SUBJECTS, MATERIALS, AND PROCEDURES	65
	The Subjects	66
	Materials	68
	Procedures	79
	Statistical Methods	80
IV.	ANALYSIS OF THE RESULTS OF THE INVESTIGATION	85
	Correlation Analysis	85
	Multiple Discriminant Analysis	97
V.	SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	111
	Summary	111
	Conclusions	115
	Recommendations	116
BIBLIOGRAPHY		118
APPENDIX		131

LIST OF TABLES

TABLES		PAGE
1.	Means and Standard Deviations of Kindergarten Predictors: Samples for Grades 1-4	86
2.	Means and Standard Deviations of Criterion Variables: Grade Equivalent Scores	86
3.	Correlations between Predictors and Criteria .	87
4.	Intercorrelations of Four Predictors	88
5.	Intercorrelations of Nine Predictors	89
6.	Multiple Correlations and Contributions of Four Variables	91
7.	Multiple Correlations and Contributions of Nine Variables	92
8.	R, Constants, and b Coefficients for Four-Variable Correlation	94
9.	Correlations, Constants, and b Coefficients for Two Major Predictors of Four	95
10.	R, Constants, and b Coefficients for Nine-Variable Correlations	96
11.	Comparison of R's Using Four and Nine Variables	97
12.	Scores of Achievement Groups	98
13.	Wilks' Lambda Test of Equality of Centroids .	99
14.	Group Means and F-ratios: Grade 1 Sample . . .	100
15.	Group Means and F-ratios: Grade 2 Sample . . .	100
16.	Group Means and F-ratios: Grade 3 Sample . . .	101
17.	Group Means and F-ratios: Grade 4 Sample . . .	101

18. Discriminant Analysis Using Four Variables . . 102

19. Discriminant Analysis Using Nine Variables . . 102

20. Correlations of Discriminant Functions with Predictors 104

21. Scaled Vectors 109

LIST OF FIGURES

FIGURES		PAGE
1.	Centroids of discriminant functions when both were significant in nine-variable analysis for Grade 3	107
2.	Centroids of discriminant functions when both were significant in nine-variable analysis for Grade 4	108

CHAPTER I

Introduction

Reading is generally accepted as the most important learning tool for academic success. According to one reading authority:

> Through mastery of the power to read access is given to sources of knowledge and delight, to records of the past and visions of the future. Without this ability it would be impossible to move easily and live actively in a modern community where instructions, information and guidance even in the simplest and most elemental forms of behavior are conveyed in some form of print (Inglis, 1948, p. 6).

The more immediate community of the school can be a source of failure and disillusion to the child who does not have at his command basic reading skills. In evaluating compensatory programs for the disadvantaged, reading achievement is used as a criterion. Overall surveys of schools or districts have been concerned primarily with level of achievement in reading, and evaluations of individual children with learning difficulties have usually begun with an assessment of reading ability. Problems in classroom behavior and organization were also considered allied to reading disability (Bond & Bond, 1945).

Intellectually gifted children as well as those of limited ability have experienced difficulty in learning to read. No simple relationship between reading and a single factor, such as intellectual ability, has been found (Durkin, 1966, Kottmeyer, 1947; and Witty & Kopel,

1939). Reading appears to be related to the interaction of a number of factors, the dimensions of which are as yet unclear to educators and psychologists.

Various teaching methods have been, and are still being, used in attempts to find a single method which will insure successful reading achievement for every pupil. These efforts have been described:

> For over a decade almost every basic issue in beginning reading instruction -- how to begin, when to begin, what instructional materials to use, how to organize classes for instruction -- has been debated with intense heat and considerable rancor. Laymen and self-styled reading specialists have confidently provided answers in a stream of popular books and magazines and newspaper articles. Most of these answers have been rejected with equal confidence by teachers, administrators, and reading specialists in the professional educational literature. Each side has claimed that it knows how to give our children "the best" in reading instruction (Chall, 1967, p. 1).

Schools have tried basal readers of various types and emphases, individualized reading, phonics approaches, the look-and-say method, linguistic approaches, the initial teaching alphabet and other simplified alphabets, the experience method, etc., but the fact is that despite years of effort and controversy, no method has been found which is able to guarantee success for all. Bond (1966), who coordinated 27 USOE first-grade reading studies, concluded that no one program of reading instruction was sufficiently superior for him to recommend its exclusive use.

Another approach, perhaps more fruitful in relation

to the problem of reading success and failure, is to attempt to predict those children who will encounter difficulty in reading, irrespective of the method of instruction.

Thirty-five years ago Castner discussed the need for early prediction of reading disability:

> One of the unfortunate aspects of the problem of the so-called specific reading disability is the relative lateness of the time at which it commonly comes to clinical attention. Most cases are not referred for guidance until the child has been in school for several years, at which time the problem has become so complicated by emotional factors that the child is as apt to be referred on the basis of a conduct difficulty or general maladjustment as because of any question as to his school progress or ability. Years of failure, discouragement, misunderstanding, frequently accompanied by scolding, ridicule, and even treatment as a mental defective, have had the effects that could have been expected upon personality and adjustment. The reading handicap itself is more difficult to overcome than if it had been attacked earlier, the degree of retardation is greater, there are strongly established bad reading habits to overcome as well as proper ones to build up, confidence must be restored and more favorable attitudes developed. Moreover, the effect of the reading handicaps on other subjects, such as arithmetic, often makes it necessary to extend remedial tutoring into these other fields. It is obviously desirable, then, to discover such handicapped children in an early stage in their school careers and to undertake preventive and corrective guidance before the problem takes on all these complications (Castner, 1935, p. 375).

More recently, the need for predictive measures was cited by one authority:

> An accurate group screening device which can be used to identify potential failures, who may thus be given special attention from the beginning, should be extremely useful in reducing the frequency and severity of reading disability (Harris, 1968, p. 182).

The history of prediction of reading achievement is replete with contradictory and unsubstantiated findings. A review of studies in this area revealed such weaknesses as small and limited samples, inadequate statistical treatment, and conclusions which were not based on empirical evidence.

Success in reading has been found to be related in small measure to intelligence, sex, socioeconomic status, perceptual and motor abilities, and social and psychological factors. Obtained relationships, however, have not been sufficient to account for the greater part of the variance by any one predictor. There is a need for data related to prediction based on a sizeable sample using sophisticated statistical techniques for a number of variables.

Once children with a potential for failure have been identified, an examination of their skills and abilities can be undertaken, followed by specific instruction in an attempt to prevent the academic failure which so often has marked children for their entire lives. If no one teaching method has been found to prevent reading failure, it is possible that abilities of the individual child when he enters school might offer evidence of his future success or failure.

Statement of the Problem

This study was proposed as a longitudinal study of 700 children from kindergarten through fourth grade to

determine whether measures that are readily available in most school districts, taken in kindergarten by classroom teachers, can predict reading achievement, as measured by standardized tests, in succeeding grades.

In kindergarten the children were given: the Metropolitan Readiness Test (MRT) with its subtests in Word Meaning, Listening, Matching, Alphabet, Numbers, and Copying, including the Goodenough Draw-a-Man Test (DAMT); a ranking by the teacher (TR); and a rating on a composite behavior rating scale (BRS) which included motor and speech behavior, social behavior, emotional behavior, intellectual abilities and behavior, and adjustment to the classroom.

Multivariate analysis was used to answer the following questions:

1. What is the extent to which selected kindergarten measures can predict reading achievement in grades one, two, three, and four?

2. Are these measures equally effective in predicting achievement on the four grade levels, or are different measures better predictors of reading achievement in different grades?

3. Do correlations among the measures indicate that duplication of measurement exists, or are the measures sufficiently independent to warrant their inclusion in the battery?

4. Which composite of kindergarten measures best discriminates the high, average, and low achieving readers at each grade level?

Definition of Terms

Since discriminant function requires separation into definite, even though arbitrary, groups, the following mutually exclusive categories have been used in this analysis:

Low achieving readers - those children with reading achievement scores on the MRT in the first quartile (i.e. lowest 25%) of the grade sample.

Average readers - children with achievement scores in the middle 50% of the group.

High achieving readers - children with reading scores above the fourth quartile (i.e. upper 25%) of the grade.

These children, naturally, might be categorized differently in each grade for which their scores were available.

Achievement scores were apparently normally distributed so that findings of this study may be considered applicable to other samples with normally distributed scores.

Significance of the Problem

The area of prediction of academic success at the elementary level is one with which schools are presently

concerned. The team approach, using health care specialists, psychologists, speech and hearing specialists, social workers, and/or reading specialists, is at the experimental stage in a number of school districts. Ways are being sought to identify factors related to later achievement, in reading particularly, in children at an early stage, generally kindergarten. There has been little agreement about what factors have a relationship to achievement, and further, how to measure them. Consequently, considerable effort and money have been expended without concomitant improvement in reading performance.

Assuming successful identification of potentially successful and failing students, the cost of such a complex screening process is, at the present time, beyond the ability of many school districts to support universally, and administrators have expressed a need for gross screening in the early grades, that funds may be conserved for finer screening as it becomes necessary for individual children.

The predictors used in this study have the advantage of being available to all school districts regardless of budgetary considerations. The use of some standardized measure of reading readiness for placement in first grade is widespread, so that no additional expenditure is generally required. Teachers with minimal training are qualified to administer and score such tests. The other

measures involved teacher ranking and rating on a scale which can be constructed or adapted from others in existence. Thus school districts lacking extensive staffs or funds could use this procedure for predicting which children will have difficulty.

The aim of this procedure is to enable schools to take preventive rather than remedial action after children have experienced failure. It allows for consideration of each child as an individual, since once a child has been classified as possessing the potential for failure, a program of diagnosis and instruction can be planned for him based on his individual needs.

This procedure is not proposed to replace the finer diagnosis which can be effected by teams of specialists. It is suggested as a preliminary step to identify "high risk" children who will receive the specific evaluation which may not be required or practical for all children.

Most prediction studies have been confined to a single predictor variable and/or prediction for first or second grade only. While statistically significant correlations have been found, the highest generally have been in the range of .40 and .50, with resulting limited value as the sole predictor. A measure of the effectiveness of combinations of these variables as predictors is needed. In addition, longer range prediction is desirable.

A multiple correlation and regression analysis, using the MRT, the DAMT, TR, and BRS, might well afford

more precise prediction than any other method and revealed the contribution of each variable to the total prediction at each grade level through fourth. Intercorrelations might well indicate whether the same factors were being measured. If there was duplication of measurement, savings could be effected by removing duplicate measures from the predictor battery, and substitutions could be made of measures considered more likely to add to predictive effectiveness.

Discriminant function analysis could be expected to reveal the variables best discriminating three levels of readers -- high, average, and low -- and, again, individual instructional needs of the children could be met by means of this information.

Much of the literature in the area of reading is in the realm of theory and opinion, but effective prediction depends upon precise data. The use of sophisticated statistical techniques by means of the readily available electronic computer with data for a large sample is a unique contribution of this study.

Studies of this type are certainly not original with this investigator but their seeming absence is probably related to the recency of development of computer programs. With the obvious advantages of the computer, there is no reason to depend upon theoretical concepts alone in this critical area of learning. This study, therefore, contributed in both the area and method of prediction.

Limitations of the Study

This study was limited to one kindergarten grade in the Ithaca City School District in the Finger Lakes area of New York State, followed for four successive years. Children in kindergarten during the 1964-65 school year were the subjects.

By intent, measurement of the variables was made by methods and instruments readily available in the classroom. Although their use was considered justified, these measures were not unanimously considered best for the particular purpose of the study.

CHAPTER II

Review of Related Literature

Since the 1930's, when the scientific approach and the use of measurement techniques came to the forefront in American education, studies have been conducted with the aim of predicting reading achievement (Blair & Jones, 1960). In attempting to discover predictors, pioneers in reading instruction used Pearson product-moment coefficients of correlation between a single variable, usually mental ability or the then new reading readiness tests, and reading achievement.

The early studies appear somewhat naive at this time; nevertheless the statistical technique is still a sound one. The most limiting factor, it seems from the vantage point of thirty years, was the use of only one predictor variable. Unfortunately some present-day researchers have not progressed beyond the technique of the pioneers, and have produced similar results with similar variables. Numerous replications have added little. Some experimenters have used less sophisticated techniques, depending on grouping subjects according to two variables and reporting percentage of agreement, or using chi square analysis of the frequencies.

Studies which used single variables were considered in this chapter, as well as those which used multiple predictor variables. No studies were found which used

discriminant analysis in prediction of this type.

The studies reviewed were grouped according to the predictor variable used singly: readiness as measured by any standardized test and as measured specifically by the Metropolitan Readiness Test (MRT); mental ability as measured by standardized tests, drawing tests, and specifically the Draw-a-Man Test (DAMT) used in this study; kindergarten teacher rating or judgment of a child's readiness or chance of success in first grade; behavior rating scales generally. The scale used in Ithaca in this study was not, to the knowledge of the Ithaca administration, used elsewhere and its use was not reported in any study.

Some of the investigators studied more than one predictor variable. If these variables were treated singly, they were reviewed in more than one section. Otherwise the study was considered in the section on multivariate studies.

Readiness Tests

The predictive validity of readiness tests has generally been found to be in the range of coefficients of .40 to .60 (Carr & Michaels, 1941; Weintraub, 1967). Durkin (1968) believed that reading readiness tests were the result of "disenchantment" with the ability of intelligence tests to predict, and Gates, Bond, and Russell (1939), themselves involved in the theoretical development and construction of readiness tests, stated that "It should be noted that among the tests of little or

no predictive value are many tests and ratings widely recommended in books and articles on reading readiness testing and teaching."

While the content of readiness tests varies, the tests generally measure ability to recognize similarities and differences in letter forms, to match words, to select pictures of things or actions which rhyme with words spoken by the test administrator, and to follow directions (Blair & Jones, 1960).

The Harrison Stroud Reading Readiness Profile, administered to British subjects at five years four months, was found to have correlated .592 with reading progress on the Southgate Group Reading Test at age six and .596 at age six years four months (Thackray, 1965).

Bagford (1968) used the same readiness measure in first grade and correlated scores with those of the Vocabulary and Comprehension subtests of the Iowa Tests of Basic Skills in intermediate grades. Obtained correlations were .72, .67, and .69 for Vocabulary in grades four, five, and six, respectively, and .57, .55, and .55 for Comprehension in the same grades.

The Lee-Clark Reading Readiness Test was correlated by Petty (1939) with reading grades, rather than with a standardized measure of achievement, with a resulting coefficient of .44. A more reliable criterion might have yielded a higher or more reliable r. Lee, Clark, and Lee (1934) used their readiness test as a predictor of achieve-

ment on the Lee-Clark Reading Test at the end of first grade and obtained a correlation of .49 for 104 subjects and one of .68 for 92 children who were described as having had kindergarten training. More recently Panther (1967) found a significant correlation (.66, N= 44) between the Lee-Clark administered early in first grade with reading scores on the Metropolitan Achievement Test (MAT) at the end of first grade.

A correlation of .37 between the Lee-Clark, given in the spring of the kindergarten year, and success in early first grade was found by Matick (1963), but success was measured by first grade teacher opinion, hardly a reliable criterion. Hopkins and Sitkei (1967) used this readiness instrument as a predictor and concluded that it did "at least as well" as the California Test of Mental Maturity (CTMM) in predicting scores on the Lee-Clark Reading Test and end-of-year teacher marks. Obviously a standardized criterion would have been preferable, but their findings were consistent with those of other investigators.

The same readiness test, given to entering first graders, was correlated with beginning second grade scores on the California Reading Test (CRT), Lower Primary, and for 703 subjects a correlation of .82 was found (Powell & Parsley, 1961). The authors claimed that the test was not suitable for placement, based on their findings of

correlations of .25, .50, and .48 with achievement for high, average, and low groups, respectively, but low coefficients are to be expected with homogeneous groups.

Dobson and Hopkins (1963) found that the Lee-Clark predicted achievement on the Wide Range Achievement Test in grades one through four with correlations of .46, .40, .36, and .33 for these successive grades. They obtained correlations of .36 and .51 with the California Reading Vocabulary and Comprehension subtests, respectively, in third grade, and correlations of .34 and .33 with the same tests in the fourth grade. They concluded that predictive validity of this instrument decreased with each succeeding grade.

A long term prediction study was made by Moreau (1950) who used scores of 275 sixth grade children on the California Basic Skills Test as criteria. Correlation with the Lee-Clark, which had been administered at the beginning of first grade, was .46, which she said was only slightly lower than mental age, the best predictor. She believed that this readiness measure predicted almost as reliably for sixth grade as for first grade.

The Monroe Aptitude Test was found by its author (1935) to have a correlation of .75 with the Gray Oral Paragraph Reading Test, but such a criterion cannot be considered a comprehensive test of reading. Roslow (1940) used this test as a predictor but did not use appropriate

statistical techniques or legitimate criteria (one measure was the principal's estimate of oral reading) to reach his conclusion that those subjects with higher readiness and IQ achieved better.

The Monroe was used by Dean (1939) with 116 children, and a correlation of .41 was found. It is interesting to note, however, that when Dean obtained a partial correlation of the Monroe with achievement, with mental age (MA) and MRT scores held constant, the correlation became .20. When he correlated achievement with the Monroe, with MRT scores held constant, the coefficient was .01, showing the great degree of relationship between the two readiness tests.

Banham (1958) reported a correlation of .69 for the Science Research Associates (SRA) Readiness Test used as a predictor of the MAT.

Silberberg, Iversen, and Silberberg (1968) obtained correlations of .59 for boys and .70 for girls between the Gates Reading Readiness total score and achievement measured by the Bond-Clymer-Hoyt Developmental Reading Test.

Other investigators have constructed their own instruments or have used various subtests of existing readiness measures as predictors. These studies, of course, had the resulting limitation of questionable reliability and validity of the predictors.

The Letter Matching subtest of the Lee-Clark was found, in an early study reported by Barrett (1965), to correlate with the Gates Primary Reading Test (r= .51).

A study by Gavel (1958) used non-standardized tests as predictors and rated as best predictors ability to write letters dictated, name letters, and identify letters named. The correlation of these letter tests, administered in September, with June achievement of about 40 first graders, ranged from .22 to .60.

Wilson and his colleagues (1938) tested their theory of reading readiness: "Reading readiness is in reality reading progress: in particular, progress in the initial stages of learning to read."

They believed that "No inherent qualitative differences among these children explain differences in their progress in reading . . . (these differences) can be explained mainly in terms of learning." (1938, p. 443). They correlated subtests of the Gates Reading Diagnosis Tests with the Word Recognition subtest of the Gates Primary Reading Test at the end of first grade and obtained coefficients of from .70 to .79. Using the Stanford Achievement Test (SAT) as the criterion, correlations of from .60 to .89 were found.

Wilson and Flemming (1938) used tests of naming small letters and capital letters as predictors of end-of-year achievement, and obtained correlations of .79 and .66,

respectively, for the two tests. When IQ was held constant, these coefficients were decreased only slightly. The validity and reliability of the criteria, however, were questionable, as the experimenters used 14 tests to measure reading achievement.

The Harrison-Stroud Letter Naming subtest was found by Slobodzian (1968), through the use of percentages, to be one of the measures on which successful readers scored significantly higher.

Hillerich (1966) described Olson's unpublished study which yielded a correlation of .55 between reading achievement and knowledge of letter names and ability to write letters. He also quoted McHugh's similar unpublished result. McHugh, too, found that knowledge of either capital or lower case letter names correlated more highly with reading than did any subtest of the MRT. Specific statistics, however, were not given.

Olson (1958) cited correlation coefficients of from .22 to .64 between knowledge of letter names and reading achievement measured half way through first grade, which he claimed was higher than the correlation of MA with achievement. It must be noted, however, that his measures of reading achievement were word recognition tests only. In addition, a six months prediction is apt to be higher than one over a longer period of time.

Nicholson (1958) also mentioned that tests which

measured association with names and forms of letters showed the highest correlation with learning rate for words, in the .30's and .40's generally, with the highest .51. Again, the criterion was word knowledge alone.

The Letters and Numbers subtest of the Gates Reading Readiness Test was found to be "nearly as efficient as all subtests" in predicting achievement on the Bond-Clymer-Hoyt Developmental Reading Test by Silberberg, Iversen, and Silberberg (1967) in an investigation of 222 kindergarten children in Minneapolis. It correlated .51 for boys and .58 with girls. The authors explained that the closer a predictor variable is in context to a criterion variable, the higher the coefficient of correlation. The Letters and Numbers subtest requires skills very close to actual reading.

Kerfoot (1964) used a number of subtests of various readiness instruments in his predictions, and concluded that the readiness variables most closely correlated with reading and spelling achievement were Naming Letters and Numbers and Pattern Copying.

Weaver (1968) found the MRT Copying subtest an important predictor of achievement on the MAT in her study of 77 disadvantaged children, and concluded that visual motor skills were more closely related to reading achievement for these children than were specific language skills.

Pattern Copying was the greatest predictor, (r = .519)

of achievement on the Chicago Reading Test in a study where Goins (reported by Barrett, 1965) used 14 visual perception tests using pictures and designs. Correlations of other predictors ranged from .318 to .409 (Reversals).

Gates (1940) used seven reading readiness tests which appeared to be components in various forms of present-day readiness tests -- picture directions, word matching, word card matching, rhyming, blending, letter matching, and letter and number naming -- in a battery to predict reading achievement of 173 children in New York City schools at the end of one term and obtained a"mean" r of .706. The readiness tests were a better predictor of achievement than the Pintner Cunningham Primary Mental Ability Test (PC), whose correlation was .449. The battery Gates used correlated .760 with the PC, suggesting some commonality between readiness Tests and measures of intellectual ability.

Earlier (1936) Gates and Bond found that two readiness tests had a "fairly good correlation" with reading achievement. They described them as better predictors than the Binet IQ or MA, but gave no more specific data.

Gates, Bond, and Russell (1939) cited a study by Potter which obtained correlations between matching tasks and reading achievement of .24 to .47.

DeHirsch and others (1966) used subtests of the Gates, Word Matching and Word Rhyming, as predictors and

obtained correlations of .35 and .22, respectively, with overall reading performance (ORP) which combined the Gates Reading and Gray Oral Reading scores. With their own test of name writing and one of letter naming, they found correlations of .43 and .55 with the ORP Index. Tests of word recognition and word reproduction, which are essentially tests of the initial stages of reading, showed correlations in the .40's. Her sampling and statistical techniques, which are discussed in the section devoted to multivariate studies, posed severe limitations to acceptance of her findings.

Zaruba (1968) also concluded that letter recognition had the "greatest value" for predicting success, but failed to use legitimate statistical techniques or provide any quantitative measure of "value." Silvaroli (1964) reported that his study of 475 Long Island children showed that the single factor of letter identification (Durrell Informal Test of Upper and Lower Case Identification) was as good as all or any combination of predictors. These predictors included intelligence, socioeconomic status, and maternal n - achievement.

Dykstra (1967) cautioned against the use of subtests of readiness tests because of the insufficient reliability. He also believed, yet did not justify with data, that a single subtest was almost as good a predictor of achievement as an entire battery.

Sprigle (Sprigle & Lanier, 1967) devised his own readiness test, involving verbal comprehension, size relations, visual discrimination, reasoning, understanding of numbers, information, analogies, vocabulary, and spatial relations, which he claimed gave correlations with the MRT of .83 to .90 for three age groups, and had the advantage of requiring only eight to ten minutes for administration by a nurse. This test correlated from .63 to .66 with reading achievement, but as a new instrument required further validation. Predictive validity was based on 276 children. Using 30 randomly selected children from this sample as a basis, he claimed test-retest reliability of .96.

Much earlier (1930) Deputy developed his own test and, using 103 subjects, obtained a correlation of .70. One might speculate why the test has not survived with predictive validity of such magnitude.

A form copying test, the Bender Visual Motor Gestalt Test, which is not specifically a reading readiness test and yet requires skills among those measured in most reading readiness tests, was used by Keogh (1963) with 149 subjects. Using the Lee-Clark Reading Test as the criterion at the end of first grade, she obtained a correlation of .39. The correlation was about the same (.42) between the Bender and the first grade teacher's rating of children's achievement.

The MRT, used in this study, has been employed frequently, alone and in combination with other measures of kindergarten children, as a predictor of success in later reading.

In 1936 Wright (cited by Roslow, 1940) found a correlation of .44 between the MRT and achievement on the Gates Primary, using 208 subjects. The following year Wilson and Burke (1937) reported their use of this test as one of three predictors with 25 pupils at the Horace Mann School. Correlation of the total MRT with Gates Primary Reading Tests Types 1, 2, and 3, and with the Hildreth First Grade Test showed coefficients of correlation ranging from .26 to .60. They concluded that reading readiness tests had little predictive value, but the inconsistent method in which the study was conducted (e.g. various tests given throughout the year rather than all subtests at the same time) is a limitation of the findings of this early attempt at prediction.

A Canadian study involving 545 subjects was cited by Robinson and her colleagues (1965). Zingle and Hohol administered the MRT and found a correlation of .40 with the Marion Monroe Reading Test at the end of a year.

Mitchell (1967) used the MRT to predict scores for 7310 white and 518 Negro children in Word Reading, Paragraph Meaning, Vocabulary, and Word Studies subtests of the SAT and obtained correlations ranging from .52 - .60 for Negroes

and .56 - .59 for whites.

In an earlier study (1962) he used an entire first grade (N=1170) to predict achievement on the MAT and found correlations of .475 for Negroes and .511 for whites using the MAT as the criterion.

The MRT was used as a predictor of success in first grade by Matick (1963) with 972 suburban children, and a correlation of .56 was found. The criterion, however, was teacher opinion early in first grade, rather than some more reliable measure.

Correlations of .34 and .25 were found for the MRT and the California Achievement Test (CAT) Reading Vocabulary and Reading Comprehension subtests, respectively, by Bryan (1964) in his sample of 25.

Mayans (1966) obtained a correlation of .55, the highest of her predictor variables, with the Gates in her study of 245 subjects.

The total MRT score was found to be the best predictor ($\underline{r}$= .60) of first grade achievement on the Gates Primary Reading Test by Charry (1967). For fourth grade the correlation was .54, with the MRT still the chief predictor.

Kottmeyer (1947b) found a correlation of .46 between the MRT and the Gates, which was close to the magnitude of that for the Detroit First Grade Intelligence, which yielded a correlation of .423.

Shea (1968), using the MRT, the Lorge-Thorndike Intelligence Test (LT), and her own Test of Visual Discrimination of Words, attempted to determine which was the best predictor for 76 first grade children in Connecticut. Her unsophisticated use of statistics casts doubt on her findings, but she did report a correlation of .61 between the MRT and an unstandardized word recognition test of all words in pre-primer and primer vocabularies of the Macmillan Company series and of the Ginn and Company series of basal readers. This lack of an appropriate criterion has proved a limitation of several studies.

Bremer (1959) contended that readiness tests cannot be used to predict reading achievement with any degree of accuracy after obtaining a correlation of .40 between the MRT and achievement at the beginning of second grade, as measured by the Gray-Votaw-Rogers General Achievement Tests, Primary Test, which he used with 2069 children in Amarillo.

Karlin (1957) also concluded that "It is virtually impossible to predict from the reading readiness test score how well any child in the sample will do on the reading test (p. 322)." He used 111 first grade pupils who were selected for certain characteristics of normality. The MRT was administered in September and the Gates Primary Reading Test in May. He found a correlation of .36 which was significant at the .01 level but practically useless

for prediction purposes. When the influence of chronological age (CA) and IQ was removed, the r dropped to .25.

Kottmeyer (1947a) obtained a correlation of only .46 when he used scores of 3115 children on the MRT and the Gates Reading Test.

Dean (1939) used the MRT to predict achievement on the MAT and found a correlation of .59, better than that of the Monroe as a predictor, with its coefficient of .41. He claimed that MA was a better predictor, but the significance of the difference between a correlation of .62 (MA and achievement) and one of .59 (MRT and achievement) is doubtful. He used 119 subjects in Billings, Montana. His highest R in multiple correlations was .64, for MA and the two readiness tests correlated with achievement.

Bagford (1968) quoted Bliesmer's findings of correlations in the range of .50 and .60 between reading readiness scores and measures of early reading success. Bagford himself, in a longitudinal study, used 119 kindergarten subjects in Iowa City whose scores on MRT subtests were correlated with Iowa Tests of Basic Skills subtests of Vocabulary and Comprehension in grades four, five, and six. He found coefficients ranging from the .20's (Sentences on the MRT with Vocabulary and Comprehension) to .53 (total MRT with Vocabulary in all three grades) and with Comprehension in all three grades (.54, .51, .49). He found the Harrison-Stroud a better

predictor (up to .72) but because the Harrison-Stroud was administered in kindergarten and the MRT in first grade, the results cannot be considered comparable.

Dykstra (1967) reported a number of unpublished studies which used the MRT as a predictor. Fry obtained a correlation of .62 and Hayes one of .64 when the SAT Paragraph Meaning subtest was used as the criterion. Both studies were conducted with close to 400 subjects. Sheldon and Stauffer each used the same independent and criterion variables. Sheldon obtained a correlation of .70 for 467 subjects, and Stauffer, using about 200 children in two treatment groups (the basal reader and language experience approach) found correlations of .47 and .78, respectively. Tanyzer was also reported to have found correlations in the .50's using two reading instruction methods, with the same independent and dependent variables.

Dean (1965) used subtests of the MRT in a study of 263 mid-western children and concluded that the Numbers subtest of the MRT was "the most outstanding indicator for success or failure" in reading in grades two and three. For grade three achievement the Matching subtest was valuable, but not for grade two.

Wartenberg (1967) concluded that the Numbers subtest was the highest single predictive measure for the Word Study subtest of his criterion, the SAT. He did not

test the relationships of his predictors with an average or overall reading achievement in his study of 98 subjects.

Dykstra (1967) cited seven investigations in which the Numbers subtest was the best predictor, with correlations ranging from .43 to .71.

Hildreth and Griffiths (1939), using unique statistical treatment, claimed good predictive validity (r= .70), but Robinson and Hall, cited by Dean (1965), found a medium correlation of .51. Dean (1965) also reported Robertson's 1957 study which followed children through ninth grade. She found a significant relationship between the Numbers subtest alone of the MRT but concluded that the MRT yielded the "least significant" results of all the standardized predictors.

Kingston (1962) correlated the subtests and total score of the MRT with the subtests and total score of the SAT for 272 white fourth and fifth grade pupils as the basis for a multiple correlation. He found that the MRT, given in first grade, correlated significantly with all areas of the SAT in both third and fourth grades, but that the magnitude of the coefficients was such that prediction for individuals was not indicated. While the correlations were statistically significant, the range was .30 to .60. The Matching and Numbers subtests yielded the highest correlations. The greatest R was .390 for fourth grade girls' scores on the Paragraph Meaning.

There appear to be several limitations to Kingston's approach. First, the use of subtests along with total scores did not give a true multiple correlation, since the total score is made up of subtests. The subtests used singly or in combinations less than the total lack reliability, despite the fact that a number of studies mentioned have used them as single or independent predictors. Findings based on subtests as predictors are therefore to be questioned. In addition, multiple correlations in the range of .145 to .390, while they may be statistically significant, as were correlations in other studies reviewed in this chapter, cannot be considered useful.

Thus it may be observed that readiness tests as a whole do not differ greatly in their ability to predict achievement on various grade levels, although the coefficients of correlation appear to decline somewhat in successive grades. The MRT, one of the predictors in the present study, is as efficient as the Lee-Clark, Monroe, and Harrison Stroud, and yet the correlations are not great enough to permit the test to be used as the single predictor of reading achievement. Subtests and non-standardized tests, while yielding relatively high correlations in reported studies, do not possess the reliability necessary for predictive efficiency. According to Inglis (1948, p. 84) "The use of a reading readiness test concerned with verbal facility will not in itself discriminate

between those who will fail and those who will succeed in the mastery of elementary reading", and the studies reviewed appear to bear out his contention.

Intelligence Measures

A second generally accepted predictor of reading achievement is mental ability, as measured in a variety of ways. Strang and others (1961) theorized that there is no relationship between gains in reading comprehension and initial IQ; therefore children with low IQ's can learn to read. Durkin (1966) corroborated this opinion with her experience with preschoolers who were reading. She said that high IQ's were not always evident, and agreed with Bruner's statement that "We begin with the hypothesis that any subject can be taught effectively in some intellectually honest form to any child at any stage of development (Bruner, 1960, p. vii)."

Jackson (1944) found that 43.3% of the 300 retarded readers in his investigation had an IQ of 105 or higher, indicating to him that reliance could not be placed on intelligence alone in the area of prediction.

Bryan (1964) also observed that the bright child is not always a successful reader, and that in the early grades children with comparable intelligence scores develop at "vastly different" rates.

Yet Wilking (1941, p. 268) believed that "intelligence is generally conceded to be a definite prerequis-

ite for success in reading." He cited Ladd's study which showed correlations of from .60 to .65 between group intelligence tests and reading achievement. He indicated, without data however, that the relationship between reading and nonverbal group intelligence tests is much lower. This may be explained by Inglis, who said that most IQ tests contain verbal elements in the statement of a problem or in the response required, or in both. The correlation, therefore, is due in part to common elements in the test situation.

> If it is accepted that the mind can best be described in terms of a general factor g and certain additional group or specific factors, the correlation between reading and intelligence may be explained by the employment of common mental abilities. Obviously, g will be common to both tests, and the verbal factor will have some influence on the intelligence test unless it is devoid of verbalism in both stimulus and response (Inglis, 1948, p. 44).

Inglis believed that MA was slightly inferior to reading readiness measures in predicting reading achievement, citing correlation coefficients in the .50's and .60's as typical of intelligence measures as predictors. Weintraub (1967) believed that intelligence measures were less effective than reading readiness tests which had predictive value in the range of .40's through .60's. Kerfoot (1964) found intelligence not as good a predictor as tests of visual discrimination.

Harootunian (1966), in reviewing studies in this area, concurred. "The substantial relationship between

intelligence and reading is not unexpected. What is interesting, however, is that intelligence is not the dominant variable (Harootunian, 1966, p. 389)."

Harris believed that the relationship between intelligence and reading is low to moderate initially but increases as the child advances through school. His belief was that "As the nature of the reading task becomes more one of comprehension and interpretation, intelligence becomes a stronger determining factor (Harris, 1967, p. 341)."

Durkin (1964) found in her study of early readers a correlation of .40 between IQ and achievement at the beginning of first grade which increased to .79 at the end of fifth grade.

That correlation of reading age (i.e. achievement) with IQ increases with CA was found by Fransella (1965), who obtained partial correlations between IQ and reading age, with CA held constant, of .54 for children aged 6.2 to 9, .65 for those from 10 to 12 years, and .72 for those in the 13 to 15 age group.

Petty (1939) found a correlation of .48 between IQ and reading achievement, and Grilk, reported by Robinson (1955) obtained one of .64 with reading comprehension.

Mental age was found by Gates (1937) to be related to reading grade level in the range .34 to .62, with the

highest correlations in classes with the "best" instructors and the lowest with the "poorest" instructors. Gates and Bond (1936) found correlations of about .25, but Vilscek (1965) believed as a result of her study of 402 subjects that mental age and socioeconomic status were "powerful independent variables affecting first grade reading success."

Anderson, as reported by Hillerich (1966), used 443 kindergarten subjects and found that those with lower MA (52 to 65 months) gained as much in reading as those with higher MA (79 to 91 months). Hillerich, who believed that the effect of MA was minimal, reported that correlations were low, some as low as .00.

The Wechsler Intelligence Scale for Children (WISC) and the Stanford Binet (SB) were used by a number of investigators in their predictions. Because these are individual tests, their reliability is generally higher than group tests, and therefore a truer picture of the relationship with reading achievement might be expected.

Charry (1967) found correlations with the Gates Primary Reading Test of .50 and .52 for two first grade samples (77 and 114 subjects, respectively) using the WISC as a predictor, and obtained a correlation of .39 for fourth grade achievement. In his first grade prediction battery, the WISC was his second strongest predictor, but in fourth grade prediction the WISC did not make a

significant contribution.

Slobodzian (1968) used the WISC as a predictor in a study of 115 first grade children in New Jersey. Unfortunately, inappropriate statistical treatment gave little more information than the conclusion that successful readers have higher WISC IQ's than non-achieving readers.

Neville (1961) separated the difference between retarded and non-retarded males into scores on WISC subtests. He found that retarded readers scored significantly higher on Block Design, Picture Arrangement, Performance IQ, and significantly lower on Verbal IQ, Information, Digit Span, and Arithmetic.

The SB was found by Dean (1939) to correlate .62 with achievement in a study involving 116 subjects. It was a better predictor, he believed, than two readiness tests which gave correlations of .41 and .59 with five first grade classes in Billings, Montana. When the influence of the MRT scores was removed, a correlation of .32 between SB MA and achievement was obtained. He also reported correlations of .51 using this instrument in an investigation of 141 children by Morphett and Washburne, and .377 by Harrison with 120 subjects.

Benger (1967) used the SB as one of the predictors of the Edmonton Public Schools Word Recognition Test and the Gates Primary. She obtained a correlation of .536 for the 60 Canadian children, but since children with

high and low IQ's, difficulty in mathematics, emotional disturbances, home background problems, and physical disabilities were excluded, the validity of the prediction for a normally distributed population is questionable.

Bobbe and her colleagues (1963) correlated ability as measured by the CTMM with reading achievement on the CAT, using fourth, fifth, and sixth grade children. They obtained a correlation coefficient of .75 with the Reading Comprehension subtest. The authors did not indicate, however, when the intelligence test was administered. With this high an *r*, it might be assumed that the test was given at the intermediate grade level along with the achievement tests, since it was the investigators' hypothesis that intelligence correlated with achievement as well as did previous achievement, and therefore it was not necessary to administer achievement tests. Thus the prediction was probably not a long-range one. Matick (1963) used the CTMM scores of 972 suburban children as predictors of success in early first grade, and reported a correlation of .37. His criterion was teacher opinion, which cannot be considered as reliable as a standardized instrument.

Dykstra (1967) reported that the Pintner Cunningham was found to be correlated .75 with SAT reading by Fry. Hayes and Sheldon used the same variables and obtained correlations of .63 and .59, respectively, and Stauffer

and Tanyzer found coefficients of .43 and .49 respectively. Gates (1940) tested 173 children halfway through the first grade on the Gates Primary Reading Tests Type 1 and Type 2 and obtained a correlation of .45 with the PC.

Deputy (1930) found a higher correlation, .70, in his study of 103 subjects, and Lee, Clark and Lee (1934), using the same independence variable, obtained a correlation of .39 with the Gates Silent Reading Tests, Types 1, 2, and 3, evidently forerunners of the present-day Gates.

In her longitudinal study of 276 children, Moreau (1950) found a correlation of .39 for IQ and .53 for MA based on PC scores with achievement as measured by the California Basic Skills Test, compared with the .46 she found for readiness.

Santoro (1967) used the Otis Quick Scoring Mental Ability Test as a predictor for 21 pupils and obtained a correlation of .74 with the California Reading Test. The size of the sample may have influenced the correlation. Use of the same predictor by Gavel (1958) yielded a correlation of .45 for IQ and .44 for MA for about 40 first graders.

The Lorge Thorndike was used by a number of investigators whose resulting coefficients of correlation were similar to those of other measures of intelligence. Wartenberg (1967) did not report individual correlations in his study of 91 suburban Philadelphia children, but

included IQ as one of the predictor variables "of most significance."

Edwards and Kirby (1964) predicted from first grade intelligence scores on this instrument to third grade, and found a correlation of .46 with achievement on a subtest of the SRA, and .50 with total achievement. Their study involved 336 pupils in Kansas.

Matick (1963) obtained a correlation of .31 in his study of 972 suburban children. Correlations of .49 and .47 (for raw score and IQ, respectively) were found by Panther (1967) in his study of 44 children. Shea (1968) did not report the correlation of LT scores with reading achievement, but obtained a multiple correlation of .76, the highest of any combination, for LT scores and scores on her own Visual Discrimination Test with an unstandardized criterion.

Some early investigators used the Detroit First-Grade Intelligence Test as a predictor. A study by Lee, Clark, and Lee (1934) yielded a correlation of .41 for IQ and .40 for MA based on the use of this instrument with 92 subjects. Morphett and Washburne (1931) obtained a correlation of .59 for 141 children, and Kottmeyer (1947) found a correlation of .423, slightly lower than that for the MRT in the same study.

Thackray (1965) used the Kelvin Measurement of Ability Test for Infants in his study of 182 subjects and

obtained a correlation of .44.

Drawing tests have been purported to possess a high correlation with intelligence scores on other instruments and have therefore been used by researchers as predictors of reading achievement. Petty (1939) used Peck and Manuel's Non-Language Prediction Test for Young Children as a predictor of reading grades and obtained a correlation of .48, the same as that she found for IQ. Of course, a standardized criterion would bave been preferable.

Koppitz and others (1959) used the Human Figure Drawing Test with 1431 first grade children in Columbus and found a correlation of .46 with the MAT.

Barrett (1965) reported that a house drawing test used by Beck and Beck yielded very low correlations (under .20) with the American School Achievement Test, Primary I Battery.

Panther (1967), in his study of 44 children in a campus school, used two drawing tests, the Rutgers Drawing Test, Form A and the Goodenough-Harris, which was used in the present study as the DAMT section of the MRT, and obtained a non-significant correlation of .26 with the MAT Reading subtests and a significant r of .34 for the Goodenough. Neither is of practical value, however.

From his study which yielded a correlation of .64 for the DAMT with achievement and .33 for group intell-

igence scores with the same criterion, Easley (1964) concluded that kindergarten age drawings reflected an aspect of development more related to reading readiness than the visually structured group intelligence tests.

Shipp and Loudon (1964) also used the DAMT as a predictor of achievement for 115 first graders on the Gray-Votaw-Rogers Primary Achievement Tests. The DAMT correlated .51 with total achievement. The highest correlation with a subtest was .42 with Reading Comprehension and the lowest was .30, for Reading Vocabulary. Verbal intelligence measures have tended to predict vocabulary better than comprehension. They quoted Cronbach's statement that predictive validity rarely surpasses .60, believing that the DAMT was a good predictor.

Wilson and Burke (1937), using the DAMT of the MRT, found correlations of .30 to .48 with the Gates given at the end of the year. Wilson and Flemming (1938) found the DAMT at least as good a predictor as the SB, obtaining correlations of .372 for the former and .333 for the latter with achievement as measured by 14 tests, a questionable criterion, in their study of 25 high IQ children at the Horace Mann School.

The DAMT was used by Zaruba (1968) but her inadequate statistical treatment and use of reading grade placement (a subjective decision made by teachers)

as the criterion make it difficult to assess the relationship between this predictor and achievement. Her conclusion, after presenting various percentages, was that the DANT is a "limited" predictor of teacher evaluation. Had the SAT, which was used as an additional predictor, been used as the criterion for prediction of achievement, along with more meaningful statistical treatment, the study might have had some value.

The conclusion may be reached that while intelligence measures have some predictive value, as do readiness tests, the predictive validity is not of a magnitude sufficient for use of this variable as the sole predictor.

Teacher Judgment

The ability of teachers to predict children's academic success as a supplement to or substitute for standardized instruments has been claimed by several educators (Dyer and Beall & Holmes in Slobodzian, 1968). Anderson (1949) observed that readiness is "not difficult for the observant teacher to determine." Dykstra (1967) also believed that teachers are able, after a short time with children, to predict their reading success.

Investigations where teacher judgment has been dependent on a scale are reviewed in the section on rating scales. Studies in this section are restricted to those in which teacher judgment consists of a broad

appraisal of the child or a ranking.

Henig's (1949) study of 98 first grade children in Newark, New Jersey, resulted in a correlation coefficient of .59 between teacher judgment and first grade achievement. It must be noted, however, that achievement was measured by teacher grades, so it is possible that the principle of a self-fulfilling prophecy was operant.

Matick (1963) found a correlation of .429 between kindergarten teacher opinion and success early in first grade, also measured by first grade teacher opinion. With such a large sample, 972 suburban children, use of a standardized instrument of achievement would have produced definitive results.

Teacher ranking was used by Lee, Clark, and Lee (1934) as a predictor of success on the Lee-Clark Reading Test, with resulting correlations of .11 to .88, for small groups of 9 to 16 subjects. With such a broad range of coefficients, further research was clearly indicated and no conclusion could be considered applicable as a result of their findings. The small groups may have affected the results.

Kottmeyer (1965) claimed that teachers were better predictors than tests of readiness and intelligence, but his use of percentages only as evidence leaves doubt as to the statistical significance of his findings.

Teacher rating of general ability was used as a

predictor of achievement by Thackray (1965) on the Southgate Test, and he obtained a correlation of .51, which was somewhat higher than the .44 he found using the Kelvin Measure of Ability as a predictor.

A study by Carr and Michaels (1941) yielded an average correlation of .79 between teacher ranking early in the year and achievement on the Gates Primary Reading Test at the end of the year. An average correlation does not appear to be an appropriate statistic, but the range, from .64 to .94, indicated a substantial to high relationship (Crowley & Cohen, 1967 p. 57) between achievement and teacher ranking. This was a relatively short term prediction, however.

Teacher ranking in November was found by Wilson and Burke (1937) to be correlated .86, .76, and .85 with three Gates Primary Reading Tests administered in May of first grade. The investigators found higher predictive validity for teacher ranking than for any other measure they used -- various readiness subtests, intelligence, chronological age, drawing a man, grip, and motor coordination. Another study (Wilson & Flemming, 938) found correlations of teacher judgment and achievement in the range of .60 to .83 for a November - May prediction.

Zaruba's (1968) study of several predictor variables found that subjective appraisal by the teacher was a "useful" predictor, but her use of percentage to

reach this conclusion was clearly inappropriate. Teachers ranked children as high, average, and low. Their consideration included, but was not limited to, observation of physical development, coordination, general knowledge, concept development, verbal fluency, vocabulary, and processes of tests (as opposed to products).

Tyler (1968) followed 945 kindergarten children through fourth grade, when his sample decreased to 419, and concluded that there was a significant relationship (at the .01 level) between teacher judgment and mental maturity, memory, discrimination of sounds, forms, and colors, motor control, specific adjustment behaviors, interest in books and reading, and work habits. The relationship with reading achievement in fourth grade was being sought, and since there was no significant relationship reported it would seem that none was found.

Reading achievement was not predicted by teachers in an investigation by Pharis (1967), but rather reading readiness, which he termed kindergarten achievement. Prediction for 515 kindergarten children early in the year yielded a correlation of .591 with the MRT. Teachers rated on the basis of social-emotional factors, which had no relationship to the MRT.

The variable of teacher judgment has not been tested as frequently as the other predictor variables, readiness and intelligence. The studies that have been found

obtained correlations between .11 and .94, but the statistical and research design limitations of almost all the studies testing predictive ability of teacher judgment have cast doubt on the findings. These contradictory and indecisive findings demonstrate the need for the use of appropriate statistical techniques with sizeable samples to determine whether teacher judgment is a valid predictor.

Behavior Rating Scales

These measurements have generally required subjective teacher judgment in rating children, using a scale or guide for the teacher to rate specific behaviors rather than an overall impression of readiness or ability, as discussed in the previous section.

Johnson (1957) reported that as early as 1927 Bird found that "negative emotional states" interfered with learning in children four to six years old. She also cited Castner's belief that certain negative traits, especially excitable and unstable personalities, tended to predispose failure in reading.

Durkin (1960) concluded after studying six first grade boys in a case-study approach that:

> A variety of factors, other than that of intelligence, noticeably affect a child's progress in learning to read. Certain personality traits seemed especially significant as deterring factors: in particular, the tendency to be passive and shy and to be lacking in self-confidence (Durkin, 1960, p. 32).

It would seem then, that the recording of aspects of

emotional and social behavior might contribute to accurate prediction of reading success.

An early study in which the authors believed that teachers have the ability to "recognize those qualities, abilities, potentialities, and interests which have much to do with a pupil's success in learning to read" (Lee, Clark, & Lee, 1934) found a correlation of .49 with their standardized reading achievement test. Twenty questions were asked related to the 92 children's home environment, physical condition, ability to relate to others, interest, and ability in school.

Home environment (Thackray, 1965), as measured by a multiple choice Picture Vocabulary Test constructed by the experimenter, teacher ratings of language and speech, and notes made by the experimenter concerning the socio-economic backgrounds of the children's home environment, was found to correlate with reading achievement on the Southgate Group Reading Test administered at age six (A) and at six years, four months (B), as follows:

Predictor	A	B
Vocabulary Profile	.431	.389
Teacher rating of language	.492	.468
Teacher rating of speech	.386	.406

Teachers also rated emotional and personal attitudes of the children with the following correlations:

Predictor	A	B
Self-confidence	.337	.273
Cooperation with adults	.341	.287
Cooperation with children	.103	.134
Persistence	.361	.353
Stability	.219	.165
Prevailing attitude	.356	.332

These predictors were not explained further.

Ratings on all predictors were made when the 182 children were five years, four months of age, in their second term in British schools. Slobodzian (1968) reported Wright's 1936 study in Indiana, in which he used as predictors a pupil rating scale of personal characteristics, reading readiness, social adjustment, emotional stability and ability to learn. Teachers rated the 200 subjects from very poor to excellent on items in each category at the beginning of first grade. End-of-year teacher grades, not an ideal measure from the standpoint of reliability, were used as the criterion. He found the pupil rating scale the best predictor with correlations of .61 and .64 for the two years of his study, yet he reported predictive validity of the readiness test as .61 and .62, surely not different. At any rate, the scale appeared to be a useful predictor of a magnitude similar to that of reading readiness tests.

A five-point readiness status scale was used in an investigation by Kermoian (1962) involving 276 first grade children in San Francisco. A correlation of .77 was found between the scale and the MAT.

Alshan (1965) used teacher ratings on a five-point scale with 82 children and found that these ratings ranked second to auditory blending and consonant recognition as predictive measures.

Orear (1951) hypothesized that a relationship existed between reading achievement and social maturity as measured by the Vineland and the Munn scales, which correlated .33 with each other. She found, using the Garvey Reading Test scores of 250 California children as the criterion, that the predictive ability of the Munn was .418 and the Vineland only .216.

Medinnus (1961) developed a first grade adjustment scale with 50 items grouped into five categories: physical status and motor behavior, social behavior, emotional behavior, intellectual abilities and behavior, and adjustment to classroom membership and requirements. He reported low correlations of this scale with chronological age (.21) and IQ (.42). He believed that the inter-rater reliability, .77, was high enough to indicate "moderate agreement." He cited as advantages of the scale the facts that it was easy to use, that the language and terminology were understandable and meaningful to teachers, and that the adjectives describing the ends of the scale were not so extreme that teachers would hesitate to assign such values to children's behavior. He did not report predictive validity of his instrument, but Henderson and Long (1968) reported their use of a scale

partially derived from Medinnus'. These investigators had teachers of 192 children rate them on 24 bi-polar dimensions of classroom behavior, some of which were: "can work quietly," "is able to play in a group," "follows directions," and "talks to other children." They did not predict reading achievement, merely readiness, as measured by the MRT, and obtained a correlation of .39, which led them to conclude that

> ...it has long been recognized that teacher judgment based upon observation of classroom behavior is a sound predictor of reading readiness. The present study reaffirms this relationship and suggests the usefulness of this relatively simple and reliable scale for the assessment of social maturity in the first few weeks of grade one (Henderson & Long, 1968, p. 43).

According to Crowley and Cohen (1967, pp. 56-57), a correlation of .20 to .40 has "slight" use in predicting, making the investigators appear overconfident. This scale proved in the Henderson and Long study to have a fairly low correlation with the readiness test, suggesting that it should not be used as a substitute; however, in combination with a readiness test, it might contribute to a multiple correlation.

Ransom (1969) reviewed the First Grade Screening Test which was developed by Pate and Webb to identify children "who will probably not make satisfactory school progress without special assistance." Using a sample of 952 children, they obtained a correlation of .60 with the SAT at the end of first grade. No concurrent

validity was reported. Test items were grouped by the reviewer as follows:

Fine motor skills through production of body concepts by pencil drawing	10%
Fine motor skills through line drawing and production of figural concepts	10%
Judgments of appropriate play and social action	20%
Ability to follow minute directions through visual motor response	10%
Evidence of vocabulary and conceptual information	50%

Administration time was reported as 30 to 45 minutes, with separate forms for boys and girls. The reviewer noted that evidence of construct validity was lacking, but believed that it was a good predictor of children who were high risks for academic failure.

A teacher questionnaire devised by Mayans (1966) asked teachers to rate children on a three-point scale (above average, average, below average) in the areas of language development, self-concept, and social skills. In her study of 245 subjects, the 30 item questionnaire correlated .45 with the criterion, the Gates Primary Reading Test. This measure correlated .44 with the Binet and .54 with the MRT.

Inglis (1948), to support his belief that reading readiness tests are not the best, or only, predictors, cited a study by Robertson and Hall who found that while the median r between reading readiness and achievement was .58, the use of a teacher rating scale yielded a correlation of .68.

It appears that behavior rating scales in various forms have made a contribution to prediction of reading achievement. Correlations of this type of predictor with achievement have been found in the range of .42 to .79, similar to those yielded by the other predictors discussed. Again, such a relationship does not justify use of this variable as a single predictor, but suggests that in combination it may add to the efficiency of prediction.

Multivariate Studies

Multiple regression is the technique which immediately comes to mind when considering prediction, especially in the area of reading, where no single predictor has been found to be of such magnitude as to contribute a great proportion of the variance. Unfortunately, few studies have been conducted using this technique. Before the development of computer programs, a reluctance to use multiple regression was understandable, but since the labor involved has been reduced sharply, the accuracy of prediction afforded by this technique should be of prime consideration.

Meyers and others (1968) used regression to predict achievement on the CAT in grades four, five, and six from kindergarten test and rating data. As predictors they used the C-M Picture Vocabulary test, three tests in each category of psychomotor, perceptual spread, linguistic, and figural reasoning, and a digit span test. In addition

they used ratings on a nine-point scale of ten characteristics of test behavior. Multiple correlations for subtests of the CAT ranged from .644 to .756, with an R of .740 for total achievement. The best predictor was found to be the O-M Picture Vocabulary Test. According to the researchers, "most striking in the results is the value of behavior ratings in anticipating later achievement." The highest correlation found was .50 with the Attention category on the rating scale. Explanation of the items on the scale, unfortunately, was not made in the report. The size of the sample, 57, is a limitation of this study.

Silberberg, Iversen, and Silberberg (1967, 1968) used the multiple regression technique with IQ, CA, and the five subtest scores on the Gates Reading Readiness Tests as predictor of achievement on the Bond-Clymer-Hoyt Developmental Reading Test. They found the contribution of age negligible and observed that IQ is not a good predictor, although the multiple R was somewhat higher when they were included. An R of .59 was obtained for boys, .69 for girls, with none reported for the total group. The regression equation for boys was:

Predicted score = .39 (Letters and Numbers) +.32 (IQ) +1.2 (Word-Card Matching) - 10.73

For girls it was:

Predicted score = .43 (Letters and Numbers) +1.18 (Word Matching) +.30 (IQ) +.99 (Rhyming) - 9.41

This study, with 222 subjects, did not use the total Gates score as a predictor. Since the subtests are short, their reliability is suspect; therefore it would seem that there was little need for a multiple correlation of the subtests. A Pearson r between the total Gates and the criterion might have sufficed, since the contribution of age and IQ were relatively unimportant. The authors used subtests, however, because

> Through this type of analysis one can weigh the relative contributions of the component parts of a test used in prediction. If those portions of a test that provide redundant or only slight additional information are eliminated, the total time required for testing may be substantially reduced (Silberberg et al., 1968, p. 214).

One of the most popular recent studies (de Hirsch, Jansky, & Langford, 1966) attempted to predict reading failure for 53 children using kindergarten tests. The researchers developed a predictive index based on the number of tests on which the child scored at or above the critical score, and claimed to have identified the "overwhelming majority" of children who were later found to be failing a writing test, the Gates Reading Test, the Gray Oral Reading Test, and a spelling test at the end of second grade.

For a number of reasons, this study cannot be said to have accomplished its goals. The sample, which was small, consisted of children within 1 SD of 100 on the Stanford Binet. Parametric statistics were not used

because distributions of scores on the predictor tests were not normal. The inference is that the tests were too easy or too difficult and were therefore not appropriate. The tests themselves are questionable in terms of reliability and validity. Some were constructed by the authors, some were parts of other tests, and some were tests of skills which were difficult to measure on a continuum (i.e. tying a knot).

Multiple regression would have been appropriate for this type of study, and, according to Zieky and Page, the use of it

> . . . with its separate weighting of each predictor, not only would obviate the need for such an assumption that each test contributes equally in predictive power, but also quite probably would have increased the predictive ability of the combination of tests (Zieky & Page, 1968, p. 366).

Shea (1968) used a number of measures -- the MRT, IQ, and her own test of Visual Discrimination to predict word recognition, an unstandardized test of all words in the pre-primer and primer vocabularies of the Macmillan Company and the Ginn and Company reading series. In addition to poor selection of a criterion, when the Gates, MAT, and SAT were generally available, she did not use multiple regression but merely reported multiple correlations as follows:

Word Recognition with:	R
MRT, LT	.66
VD, MRT	.73
VD, LT	.76
VD, MRT, LT	.66

The last R, that of all predictors with the criterion, was equal to that of the LT alone. The addition of two independent variables did not contribute to the prediction. A multiple regression, however, could have been employed to indicate the relative importance of the predictors. Her use of quartiles for the criterion and comparison of placement in quartiles on the predictors with them provided no useful data of statistical significance.

Hampleman (1963) reported in a review of the literature that MA, teacher rating, and various combinations of MA, reading readiness, and teacher ratings have been found to be the best predictors.

Recent doctoral dissertations have used the multiple prediction approach which has yielded higher R's. Andras (1965) used a number of predictors -- tests of pattern copying, identical forms, auditory discrimination, phonemes, word meaning, and listening, and found the test of phonemes the best predictor of reading achievement on the SAT, with correlations of .33 to .65. The best multiple prediction (R=.417 to .664) combined phonemes and identical forms.

Mullis (1966), using the MRT, MAT, and teachers' grades in reading and arithmetic in first grade to predict fifth grade achievement in all areas, not reading alone, on the CAT, obtained R's as follows:

Sex	Original sample	Cross validation group
Boys	.657 - .851	.587 - .812
Girls	.740 - .857	.627 - .810

He did not explain reasons for a range, or why correlations were not obtained for the total group.

A sample of 115 New Jersey first level children (primary children in a nongraded organization) was used by Slobodzian (1968) in an investigation of achievement on the Gates McGinitie Reading Test. As predictors she used the WISC, Illinois Test of Psycholinguistic Abilities (ITPA), the Letter Naming subtest of the Harrison-Stroud Readiness Profile, and the Lee-Clark Reading Readiness Test. She divided subjects into successful readers and a non-achieving group (one might question why a score of 1.4 was considered non-achieving if the grade norm was 1.9 at the time of testing) and then compared scores of these two groups on each predictor by means of chi square analysis and percentage of agreement on predictor and criterion. Such comparisons gave little information about the predictive efficiency of the independent variables, the relative importance of the various predictors, or of the value of combining the predictors. Thus much of the value of the study has been lost.

Weaver (1968) used the MRT, ITPA, SB, and Peabody Picture Vocabulary Test (PPVT) to predict MAT subtests of Word Knowledge, Word Discrimination, and Reading for 77 Negroes in a program for the culturally disadvantaged.

Since total reading is the measure classroom teachers and school personnel use, it may be asked why an average or overall reading score was not used as the criterion. The optimum predictive batteries she found were:

> For Word Knowledge -- MRT Copying, SB MA, ITPA Decoding
> For Word Discrimination -- MRT Copying, SB MA
> For Reading -- MRT Total, MRT Copying, SB MA, ITPA Visual Decoding, ITPA Motor Encoding, ITPA Auditory Vocal Sequencing.

From her investigation she concluded that visual motor skills were more closely related to reading achievement in culturally deprived children than were specific language skills. This conclusion is questionable since prediction of the Reading subtest of the criterion depended on the MRT, SB, followed by the ITPA (presumably listed in order of contribution). The PPVT was not found to contribute, but all the other predictors were listed as contributors, and the MRT and SB have a large language factor. She also mentioned finding that the MRT and SB correlated significantly higher than the ITPA and PPVT with the criteria.

Wartenberg (1967) also overlooked the practical importance of an overall criterion in his investigation of 91 suburban Philadelphia children. As predictors he used the MRT, LT, the Wepman Test of Auditory Discrimination, and tests of letter identification, visual discrimination, memory span, and associative learning.

The SAT was the criterion at the end of first grade, but multiple correlations were computed for subtests only of this instrument. Thus his findings were difficult to generalize or report in a concise or useful form to the educator. He did report that the group readiness test (MRT) was the best predictor of the vocabulary subtest. Other predictors of this criterion were the vocabulary subtest of the MRT, memory span, associative learning, and IQ. Other findings were reported in an untabulated form and were difficult to group. Intelligence was found not to be the best predictor of any of the criteria, but he pointed out that intelligence in his sample was not normally distributed. Auditory discrimination, visual discrimination, memory span, and associate learning were also found not to be the best predictive of any subtest, but made some contribution. No regression equations were given which would have been helpful in individual prediction.

First- and fourth-grade achievement of 67 Pennsylvania children was predicted by Charry (1967) who used the Gates Primary Reading Test and the Gates Reading Survey as the criterion in first and fourth grades, respectively. Predictors were the MRT, WISC, Van Wagenen Readiness Scales, and the Detroit Tests of Learning Aptitude. He found the best predictors of first grade reading achievement ($\underline{R}$= .70) to be the MRT, WISC,

Van Wagenen, MRT Numbers, and Vocabulary of the WISC. The last had a negative weight. Prediction of fourth grade achievement was lower (R= .60) with the MRT, Detroit Verbal Opposites, and MRT Copying, with its negative weight, as contributors. The use of subtests and MRT Total is questionable. In the multiple prediction, the MRT is the only predictor variable useful for both grade levels.

The MRT, PPVT, Binet Vocabulary List, and a teacher questionnaire (which was a rating scale) were used by Mayans (1966) with 245 predominantly white children of three socioeconomic levels. She classified children as culturally advantaged, culturally mixed, and culturally disadvantaged, but also reported on the total group, the findings for which are discussed here. The Gates Primary Reading Test was given at the end of first grade, and all the predictors were administered early in kindergarten. An R of .60 was obtained, with the MRT the most important predictor. Although the teacher questionnaire (TQ) had a high correlation with the MRT, it contributed to the prediction. She also used discrepancy scores as predictors, which served little purpose. The multiple regression equation obtained was:

Predicted Score = 1.3 + (.03) MRT + (.013) TQ + (-.009) MB (where MB = one of the discrepancy scores)

The MRT contributed 28.39%, the TQ 9.56%, and the MRT-Binet discrepancy -2.18%. Again, the PPVT failed to contribute. The R of .60 left a large amount of variance unaccounted for, indicating that another variable or other variables should be sought.

Wright's (1936) investigation, reported by Slobodzian (1968), did not attempt multiple prediction, but studies the efficiency of a number of variables. He considered a rating scale, with correlations of .614 and .641, to be superior as a predictor to readiness scores, with correlations of .613 and .620, although they are apparently the same. Mental age had lower correlations, .489 and .547, for the two years of his study. His criterion was teacher grades, so one might question the validity of the study.

Dean (1939) obtained multiple and partial correlations with his criterion, the MAT. When MA and MRT were combined for predictive purposes, the R was .64. The Monroe Reading Aptitude Test and the MA gave a multiple correlation of .63. The MRT and the Monroe gave a multiple correlation of .59, and the correlation of all three predictors with the MAT was .64. The correlation of MA as a single predictor was .62, which he considered better than the .59 for the MRT, although there appears to be little difference. The Monroe gave a correlation coefficient of .41. His partial correlations were low,

.01 to .32, except for .54 for the MRT and Monroe with MA held constant, which would be expected. It would appear that the predictors are interrelated since none is of great magnitude when the influence of another or two others is removed.

Early researchers such as Wilson and Burke (1937), Wilson and Flemming (1938), and Wilson and others (1938) studied a number of variables but merely listed them in order of the size of the correlation. Wilson and Burke found knowledge of letters the best predictor (.59), MA next (.51), followed by matching (.49), IQ (.33), and motor coordination (-.10). Wilson and Flemming found the DAMT better than the Binet, with correlations of .372 and .333, respectively. Wilson and his colleagues, using two different criteria, found for the Gates Primary correlations of .79 for recognizing small letters, .78 for writing words, .75 for giving phonic combinations, and .70 for knowledge of letter sounds. When the SAT was the criterion, the greatest correlation was .89 for giving phonic combinations, .74 for recognizing small letters as well as for writing words, and .60 for knowledge of letter sounds. Combinations of these predictors could have contributed valuable information.

Zaruba (1968) used letter recognition, DAMT, subjective rating by teachers, and the SAT as predictors of reading grade placement. In addition to not explaining

why the SAT was not the criterion and why a subjective measurement by teachers such as reading grade placement was used as the criterion and not as a predictor, the investigator diminished the value of the study by using a complicated percentage procedure which gave no information about the relationship of the independent variables to each other as predictors. She concluded that letter recognition had the greatest value, that subjective appraisal by the teacher was useful, and that the DAMT was a limited predictor. She stated that the relationship between the SAT and teacher evaluation was the strongest, which she attributed to the brief time between the SAT, as a predictor, and assignment of grade placement scores (high, average, and low). It would seem that she overlooked the halo effect once teachers had learned the SAT scores. Her conclusions, based on 93 first grade children in California, were somehow generalized to reading readiness, although use of readiness measures was not reported. Again, a valuable opportunity was lost.

A study of 300 children in Norway (Gjessing, 1967) combined a number of predictors but did not report correlations of individual predictors. Prior to first grade the children were tested two hours a day for six days to determine mental age, verbal comprehension, auditory and visual memory and discrimination, sound blending, and numerical comprehension. School readiness was measured in

terms of social adjustment, obligation to a task, ability to concentrate, rate of work, initiative, social contact, independence, and a conception of reality. Speech therapists observed linguistic development and hand preference. The combination of reading readiness and school readiness (with specific measurement unreported and undefined) yielded an R of .57 with reading development (with no indication of the instrument) after one year and .48 after two years in school.

Thackray's (1965) study in England involved four categories of predictors -- reading readiness, mental ability, home environment, and emotional and personal attitudes. His correlations ranged from .103 (cooperation with children, in the emotional and personal attitudes category) to .596 for total score on the Harrison-Stroud Readiness Profile. Unfortunately he did not combine the independent variables in a multiple prediction but stopped after treating the predictors singly, even within the four categories.

Matick (1963) used as predictors two readiness tests, the MRT and Lee-Clark; two intelligence measures, the CTMM and LT; and teacher opinion in his study of 972 first grade children. He listed correlations in order of magnitude:

MRT	.559
Kindergarten teacher opinion	.429
CTMM	.371
Lee-Clark	.370
LT	.310

Because he used first grade teacher opinion early in first grade as the criterion, these correlations cannot be considered to represent accurately the predictive validity of the kindergarten measures. Combinations of these variables might have shown the interrelationships and the relative importance of them.

A longitudinal study by Dean (1965) used 263 midwestern children from grades one through three. She reported her findings in terms of three criteria -- Reading Vocabulary, Reading Comprehension, and total reading as measured on the CAT. Fourteen independent variables were used, including some employed in the present study. Tests were administered at the beginning of grade one. The MRT was treated as subtests only, and the Numbers subtest proved to be the best predictor of the subtests and indeed, of all the predictors. A correlation of .548 was obtained for total second grade reading and .619 for third grade. The Copying subtest and MA were next in predictive ability with coefficients of .432 and .417 for Copying in second and third grades, respectively, and .407 and .464 for MA for the same respective grades. The R for total reading for second grade was .650 and for third grade, .737. In addition to the two predictors mentioned, she found sex and general health significantly related, at the .01 level, to reading achievement. There was no mention of which combination of the 24 gave

the best prediction, or what relationship the predictors had to each other.

Benger's (1967) study involved use of multiple regression, and she found and R of .60 combining Concentration, the Frostig Developmental Test of Visual Perception, and the Wepman Test of Auditory Perception. The SB, which correlated .536 with the criterion, did not contribute to the multiple prediction. This study was severely limited by her sample, described in the section on intelligence measures.

Few of the studies which employed a number of predictor variables used an adequate statistical technique. Where multiple correlation was used, coefficients ranged from .417 to .810 with many in the .60's. It would appear that combining predictors has resulted in slightly higher correlations than those found when these variables were used singly.

CHAPTER III

The Subjects, Materials, and Procedures

The purpose of this study was to determine whether measures that are readily available in most school districts, taken in kindergarten by classroom teachers, could predict reading achievement as measured by standardized tests in succeeding grades. There were several aspects of this problem. First, it had to be established whether the individual kindergarten measures were good predictors of reading achievement on the various grade levels. In addition, their independence of each other had to be ascertained through intercorrelations of these measures. Next, it was necessary to determine whether combinations of some of these measures would predict more efficiently than a single predictor. Multiple regression equations were to be found to indicate relative importance of the predictors. In addition, in order to avert the burdensome task for teachers of computing regression equations for each child to determine his reading potential at the four grade levels, it was considered important to determine which combinations of predictors would predict a child's membership in high, average, and low reading achievement groups, with the objective of providing early assistance for those who would be expected to be inadequate readers.

The subjects, materials, and procedures, and the

statistical techniques used in arriving at answers to these aspects of the problem are described in this chapter.

The Subjects

Seven hundred children started kindergarten in 1964 in 13 public elementary schools of Ithaca, New York. At that time there were approximately 7,800 children in grades K-12.

The Ithaca City School District comprises about half of Tompkins County. It included, two junior high schools, and 13 elementary schools. Three of the latter are located in downtown Ithaca and the remaining 10 are in other areas of the city, or in the suburban and rural areas they serve. About 90% of the children in the district attend public schools. Approximately 90% of the children entering the public schools finish high school, and over 60% of those who graduate continue their education (League of Women Voters of Tompkins County, 1967).

The school population is considered by one of its chief administrators to be normally distributed with respect to IQ and socioeconomic status, variables which have been found to be related to school achievement. Four of the elementary schools, Belle Sherman, Cayuga Heights, East Hill, and Northeast, were reported to be the "silk stocking", or high socioeconomic status and high ability schools. She reported Caroline, Fall Creek,

Glenwood, South Hill and West Hill to be average schools. The schools described as having lowest achievement and socioeconomic status are Central, Danby, Enfield, and Henry St. John. The last group consists of two inner city schools and two rural schools. Included in the average and high groups are city, suburban, and rural schools. Occupations in the district range from farmers and inner city unskilled laborers through skilled workers in the several factories, white collar workers in the downtown area, to students and faculty at Cornell and Ithaca College, and bankers and company presidents.

The fact that specific schools represent specific strata of Ithaca society has not presented any difficulty in this study. The entire kindergarten class of 1964 has been treated as one group without reference to the schools children attended at any time through fourth grade, 1968-1969.

Although approximately 700 children began kindergarten in 1964, those for whom there was complete kindergarten data as well as at least one set of reading scores in grades one through four numbered 553. Of this number, not all were represented at each grade level. There were 478 first grade scores, 440 for second grade, 373 in third, and 368 fourth grade scores. A group of about 27 children in one school learned to read using ITA and were not tested at the end of first grade. They had

all made the transition to traditional orthography in second grade, so their scores were included in data for second, third, and fourth grades.

In some cases, a child attended kindergarten and several consecutive grades, then moved out of the district. Some children changed schools within the district and while generally data in several schools were available, occasionally it was not possible to find scores. Some children were absent at the time of testing during one year. Another cause of missing scores was the sabbatical leave of an academic parent from one of the colleges. The nature of the missing scores, then, has not apparently influenced the normality of the distribution.

The promotion policy in the Ithaca schools is one of almost total promotion. In the few cases where a child was retained in grade, his scores were used until the year he repeated the grade. Thus low achievers have not been eliminated from the sample unnecessarily.

Materials

Ausubel (1959, p. 247) defined readiness as "the adequacy of existing capacity in relation to the demands of a given learning task."

Similarly, it has been said that "a child who possesses readiness has in his past experience and physical, mental, and emotional makeup the elements associated with success in the task to be undertaken (Whipple, 1967, p. 80)."

Leton believed that there were variations in the meaning of readiness.

> First, readiness may be viewed as the composite level of maturation in sensory and neural systems which is prerequisite to the perception and discrimination of word forms. Second, it may be viewed as a composite of mental abilities and subabilities which are required for the recognition and comprehension of reading materials. Third, it is regarded as a specific program of kindergarten instruction related to the discerning of similarities and differences in shapes, forms, and pictures, and to the recognition of sequence in a series of pictures. Finally, it may be regarded as the heterogeneous total of of all preschool activities and kindergarten experiences which enable children to assimilate vocabulary and comprehension, and to develop interests and attitudes which predispose them to successful reading (Leton, 1963, p. 915).

He explained that the first two definitions were related to characteristics of the pupil and the second two to experience and instruction.

Measurement of the readiness factors mentioned has been generally accomplished by the use of standardized readiness tests in kindergarten or first grade, although some tests of individual abilities have been used. The readiness test is an attempt to measure all, or as many as are measurable, skills by one instrument.

The Metropolitan Readiness Test (MRT), one of four predictors used in this investigation, includes, according to its authors, the following components of readiness:

> comprehension and use of oral language, visual perception and discrimination, auditory discrimination, richness of verbal concepts,

> general mental ability -- capacity to infer and to reason, knowledge of numerical and quantitative relationships, sensory-motor abilities of the kind required in handwriting, writing of numerals, drawing, adequate attentiveness -- ability to sit quietly, to listen, and to follow directions (Hildreth, Griffiths, & Mc Gauvran, 1965, p. 11).

These are measured by six subtests. The test of Word Meaning is intended as "a measure of the child's store of verbal concepts." The pupil selects from three pictures in each of 16 items the one that illustrates the word the examiner names. The Listening subtest also used 16 items to test ability to comprehend phrases and sentences instead of individual words. The pupil selects from three pictures the one which shows a situation or event the examiner has described. Mathching is a 14-item test of visual perception measuring recognition of similar pictures. The Alphabet subtest is a 16-item letter recognition test in which the child must choose from four the one letter named by the examiner. Numbers is a 26-item test of number concepts, number knowledge, ability to manipulate quantitative relationships, recognition of and ability to product number symbols, and related knowledge, such as concepts of money. This subtest, according to the authors (Hildreth et al, 1965), ". . . has repeatedly been shown to be the single most powerful predictive subtest of the earlier forms of the MRT." The last subtest is Copying, in which the child is asked to copy 16 forms, letters, numbers, shapes, to

demonstrate a combination of visual perception and motor control similar to that which is called for in learning handwriting.

Total working time for the test is approximately 60 minutes, and it is suggested by the publisher that the tests be administered in three separate sessions. This recommendation was followed when the tests were given to the subjects in this study.

Predictive validity of .54 to .73 was claimed by its authors (Hildreth et al, 1965) with the Metropolitan Achievement Test (MAT) at the end of first grade. With the Stanford Achievement Test (SAT) subtests for a smaller sample, coefficients of correlation ranged from .52 to .75. With other standardized achievement tests, predictive validity coefficients ranged from .40 (Robinson et al, 1965) to .82 (Powell & Parsley, 1969). Generally the correlations were not as high as those claimed by the publishers, but were in the range of .40 through the upper .60's and low .70's, as discussed in the previous chapter.

Reviewers of this instrument have generally been satisfied with its content validity. Osburn (1940) considered the earlier forms "emphatically worth while" despite some imperfections. The newer forms were reviewed by Griffith (1949) whose opinion was that it was "among the superior readiness tests now available."

The standardization group for this test in 1964 totaled 12,231 children in 12 states. One limitation, pointed out by the authors (Hildreth et al, 1965) was the higher than national average socioeconomic status of the group, despite attempts to obtain a representative population.

Reliability of the MRT was reported by the authors as high. The SE_M of the total score in three samples of 167, 173, and 200 was 4.2, 4.3, and 4.3, respectively.

Based on correlation of odd-even questions, reliability of the total test for the three groups was .91, .91, and .94. Reliability of the subtests, as would be expected, was lower. The Listening subtest coefficients were .50, .33, and .33, the lowest. The other subtests, however, had coefficients ranging from .58 (Word Meaning) to .89 (Alphabet). Test users were advised by the publishers not to attach significance to subtests for individuals because of the lower reliability which normally results from short tests.

The tests are scored by comparing the pupils' responses with the correct answers given on the scoring key. In several subtests there is not one correct way of responding. For example, in the Numbers subtest, the child is to put X's on three muffins. The scoring instructions appear adequate in this situation, showing the first three muffins in the row with X's and the

directions "or a mark on ANY 3 muffins, or a continuous line through, or boundary about, ANY 3 muffins." The Copying subtest has criteria for each item. In scoring one shape the directions read "The lines should be approximately the same length and intersect at about where they intersect in the model. An X is incorrect."

Raw scores are reported for each subtest and are added to product the total score. Both of these were used in this investigation.

The Draw-a-Man Test (DAMT) is an optional test of the MRT, with the back cover of the booklet left blank for this purpose. Children are simply asked to draw a man in the space provided. This test as developed by Goodenough and Harris involves intricate scoring, but as used with the MRT drawings are categorized A through E with the letters representing superior, above average, average, below average, and immature drawings, respectively. The immature category is assigned when the figure is not recognizable as a human being and parts of the body, if drawn, are not connected. Category D is assigned if most of these are included: arms, legs, trunk, head, mouth, nose, hair. In addition to these, fingers, ears, and nostrils and better proportions among parts than in D are required to be graded C. For B classification, arms and legs should be in two dimensions, there should be better proportions (trunk longer than width) and

clothing must be clearly indicated. The Superior rating is given to those drawings which, in addition to the characteristics required for D, C, and B, possess most of the following: non-transparent clothing, lines firm and meeting at proper points, and such details as neck, hands, shoulders, correct number of fingers, and a waistline.

The basis for this non-verbal measure of intelligence, as expressed by Harris (1963), was the belief that intelligence is not a unitary factor and should be considered conceptual maturity, or the ability to perceive, abstract, and to generalize. The Harris revision of the Goodenough, and the original test itself, measured, he believed, concept formation in the young child. Correlations of this test with other measures of intelligence, the Test of Primary Mental Abilities (PMA), the Stanford Binet (SB), and the Wechsler Intelligence Scale for Children (WISC), were reported by Harris:

PMA	.41	(Ansbacher, 1952)
SB	.36	(Rottersman, 1950)
	.41	(McHugh, 1945)
	.50	(Havighurst & Janke, 1944)
WISC	.47	(Rottersman, 1950)

Most of these coefficients of correlation may be considered generally useful for group prediction (Crowley & Cohen, 1967, p. 56).

Danford (1965) correlated IQ's of 107 subjects as measured by both the DAMT and the SB, and obtained coefficients of .86 for the pilot group and .71 for the

cross-validation group. He described this relationship as "comparable with SB correlations with the WISC."

Much lower correlations with the SB, .37 and .40 were found by McHugh (1945) who tested 83 kindergarten children before and after kindergarten, respectively.

Stewart (1953) also commented on the good correlation with the SB. She believed that the norms should have been updated but considered it a good test.

Yule and his colleagues (1967) randomly selected 131 nine and a half and ten and a half year old children from 2,200 on the Isle of Wight and found low correlations of the DAMT with the WISC. Correlations for three raters were .33, .34, and .41.

Predictive validity of the DAMT has been discussed in the previous chapter.

Smith found this test reliable for 2,600 children age groups from 6 to 15-16 years of age. All except the last had correlations for two administrations in the .90's. The correlation for the 15-16 group was .84. He concluded that a "drawing test probably measures somewhat specialized abilities rather than general intelligence of the conventional linguistic type (Smith, 1937, p. 761)."

The same year McCarthy (1937) reported reliability data she obtained from administering the test to 386 children in grades three and four on two occasions a week apart. They were scored independently three times, twice

by the same examiners and once by a different one. She found a correlation of .94 when the drawing was scored by the same examiner. Average inter-scorer reliability was .90. Two different drawings by the same child, as scored by the same examiner, produced a much lower correlation, .68.

Robinson (1966) found test-retest reliability of the Goodenough for three socioeconomic groups, advantaged, average, and disadvantaged .866, .844, and .843, respectively. The samples were small, 30, 51, and 40.

Two years later he reported reliability data for the three groups for grades one, two, and three (Robinson & Hanson, 1968). Coefficients ranged from .72 (average children in third grade) to .92 (advantaged children in first grade) with no apparent pattern.

Yule (1967) found fair reliability for three raters, .91, .87, and .86, and Danford (1965) reported scoring reliability of .90.

Brill (1935) tested the reliability of the DAMT and that of an abbreviated scoring method. This method involved finer discriminations, however, than the simple five-category classification used in conjunction with the MRT. He administered the test three times to 65 to 93 feebleminded boys, surely not a representative sample, and found correlations of between .68 and .80. Using the abbreviated method, correlations among the three scores ranged from .65 to .75. Correlations between

original scores and abbreviated scores for the three administrations were .95, .98, and .92, leading him to conclude that the abbreviated scoring method appeared to be a valid and reliable measure of intelligence or whatever the original Goodenough score measured.

Both the MRT and the DAMT were administered in this study to small groups of children by the kindergarten teachers or psychologists.

The behavior rating scale (BRS) used in this investigation is one developed by the Department of Psychological Services on the Ithaca City School District, adapting and combining scales which had been used by other school districts for the same purpose. Children were rated from 1 (poor) to 5 (excellent) on 30 items in five categories; motor and speech behavior, social behavior, emotional behavior, intellectual abilities and behavior, and adjustment to the classroom. The scale, including rating directions to the teachers, is in the Appendix. No reliability data has been sought by the district for this scale. Predictive validity of similar scales has been discussed in Chapter Two along with some meager information about reliability of a few.

Their teachers ranked kindergarten children according to the following instructions from the Elementary Coordinator:

> List the children beginning at 1, assigning number 1 to the most ready and numbering each to the least ready for 1st grade. In other words, if you

have 27 children your best student will rank 1, your most immature will rank 27.

The Elementary Coordinator, by design, requested the teacher ranking (TR) and BRS scores before the MRT was administered.

Near the end of first grade, in April, 1966, the MAT Primary I Battery was administered. Teachers in the district were familiar with and most were experienced in administering the test. Publishers' directions for administration were followed. The three subtests, Word Knowledge, Word Discrimination, and Reading were averaged to obtain the first grade criterion.

The MAT has been used widely in school districts. McKim (1953) objected to the use of it for diagnosis, but viewed it as valuable in measuring reading achievement. Hobson (1953) believed that the narrow range made it a more valid measure of achievement for a given grade. He commended the scaling and the manual.

At the end of each of the next three consecutive years, SAT was used to measure reading achievement. The MAT was taken off the market at that time for research and improvement, and the SAT was offered as the achievement battery by the publishers. This test was given in April of 1967, 1968, and 1969. The Primary Battery II and the Intermediate Battery I were used for grades two, three, and four. Primary I contains subtests of Word Reading, Paragraph Meaning, and Vocabulary, which

were averaged to obtain a second grade reading score. The third grade average reading score consisted of the subtests of Word Meaning and Paragraph Meaning, as did the fourth grade average reading score.

The SAT standardization program involved more than 850,000 children in 264 school systems drawn from 50 states. Different size and type school systems were represented in the sample. Generally recommended procedures for test construction and item analysis were apparently followed by the publishers (Kelley, Madden, Gardner, & Rudman, 1965).

Robinson (1959) noted that despite a few limitations, chiefly the fact that it is timed whereas the Gates test of reading imposes no such penalty, it is among the best survey tests on the elementary level, "a dependable gross measure of reading achievement."

Procedures

This study was made possible with the permission and help of the Coordinator of Curriculum, K-12, of the Ithaca City School District. The investigator was given free access to pupil, school, and district records.

Kindergarten data were recorded at the end of the year by the teachers on forms which were kept by the coordinator. The predictor scores were obtained from these sheets. Achievement scores for first through third grades were obtained from records which were also taken by the Coordinator after testing and kept in

her files. Fourth grade scores were taken by the investigator directly from computer printouts of scores.

Because all teachers did not follow instructions related to record keeping, it was not possible to locate raw scores of achievement tests for all subjects. Rather than curtail the size of the sample sharply, grade equivalent scores, which were kept for all who took the tests, were used. It was recognized that the precision afforded by the use of raw scores was sacrificed somewhat, since a number of raw score points can be represented by one grade equivalent score.

Where there were kindergarten data and at least one criterion score, a child was included in the sample. Data were compiled by the investigator on sheets for keypunching. A computer program for the IBM 360/40 was used to average subtest scores of the criteria, to add subtest scores of the MRT, and to transform teacher ranking into a score to be used as a predictor. The child's rank was inverted (e.g. child number 1 in a class of 28 became 28) and his rank was divided by the number of children in the class. A new data deck was produced by the computer with these data in concise form for statistical treatment by the computer.

Statistical Methods

Two techniques formed the basis for analysis of the data. A multiple correlation and regression program produced correlations of individual predictors (with the

MRT treated as a single predictor score in one pass and as six subtest scores in another) with an individual criterion. Using the MRT Total score, there were four predictors for each of the four criteria (end-of-grade measures). When the six MRT subtest scores were used, without a total, there were nine predictors, including the DAMT, BRS, and TR. These two separate groups of predictors were used four times each, namely to predict the average reading score of each grade, one through four.

Intercorrelations of the predictors were also obtained. This step was necessary at each grade level because different subjects were included in first, second, third, and fourth grade samples, as explained earlier. In multiple correlation, predictors which have a high degree of relationship with the criterion and low relationship with each other are sought. These intercorrelations are accounted for, but not necessarily revealed, in the multiple R. They were reported so that duplication of measurement would be obvious because in this study one of the objectives was simplification of screening. If one instrument could be used instead of several, it should be noted.

Multiple correlations were obtained showing which combinations of variables produced the best prediction, as well as regression coefficients to demonstrate relative weights of predictors because it was assumed that

achievement, the dependent variable, was associated with more than one predictor. In the regression equation, the coefficient is the multiplying constant, or weight, for each predictor contributing to the multiple R. The multiple R:

> is subject to the same kinds of interpretation, as to size and importance, as were described for a simple r. One kind of interpretation is in terms of R^2, which we call the coefficient of multiple determination. This tells us the proportion of variance in X_1 that is dependent upon, associated with, or predicted by X_3 and X_4 combined with the regression weights used (Guilford, 1956, p. 399).

The second technique, discriminant function analysis, was used to predict the achievement group (high, average, or low) to which a child would probably belong. "The conditional probability of membership in a ... pattern, given a person's measurement profile, is a known function of the person's generalized distance from the centroid for that ... pattern (Cooley and Lohnes, 1968, p. 5-3)."

According to Bock (1966) discriminant function, a weighted sum of the original variables which may be used to assign a new subject to a group to which he is most similar in terms of his scores on these variables, can be of great utility in "practical educational decisions." Anderson (1966) described discriminant function with g groups as similar to canonical correlation, a linear function of t variates:

$$y = c_1x_1 + c_2x_2 \ldots c_tx_t$$

such that the ratio of variance among groups to variance within groups is maximized. Guilford (1956) explained that the F ratio would be a mixiumum when overlapping of the distributions is minimal.

In this procedure, the Wilks' lambda criterion to test the equality of centroids is used to determine whether the separation of the groups was significant. If there is a non-chance difference, discriminant analysis may be applied to examine the group differences.

The choice of number of discriminant functions to be used was explained by Cooley and Lohnes:

> Discriminant analysis is a procedure for estimating position of an individual on a line that best separates classes or groups. The estimated position is obtained as a linear function of the individual's m test scores. Since one "best" line may not exhaust the predictive power of the test battery in distinguishing among the classes, additional discriminant functions, all mutually orthogonal, may be fitted. The maximum number of discriminants is indicated by the lesser of the two numbers g-1 and m (Cooley and Lohnes, 1962, p. 116).

In this investigation there were nine or four predictors, depending upon whether the MRT subtests or total score was used, and three achievement groups - high, average, and low. Therefore g-1 or two discriminant functions were used.

Achievement groups were determined by using quartiles. Q_1 was the upper limit of the low achieving group, Q_4 was the lower limit of the high achieving group, and the middle 50% was considered average. The low achieving

group was not necessarily considered failing, but was merely categorized as the lowest quarter of the class, those who would require further scrutiny. One advantage of using this method is that it is easily understood by classroom teachers and can be used by them to classify children without knowledge of or experience with statistics.

CHAPTER IV

Analysis of the Results of the Investigation

The problem under investigation was that of determining whether measures that were readily available to kindergarten teachers could be used to predict later reading achievement. Five hundred fifty-three children were followed through fourth grade and their reading achievement measured in order to discover whether prediction was possible.

Correlation Analysis

The first step in predicting reading achievement using the kindergarten measures was to obtain Pearson product-moment coefficients of correlation between the independent variables and the criteria, the reading achievement scores in the first four grades. The predictor variables were used twice, once with the total Metropolitan Readiness Test (MRT) score along with the Draw-a-Man Test (DAMT), the behavior rating scale score (BRS), and the teacher ranking (TR). In a separate computation, the six MRT subtest scores were used without the total test score, in combination with the DAMT, BRS, and TR. Table 1 shows the means and standard deviations on the predictor variables of the sample of children used for each grade. An examination of these statistics reveals that the mean scores of the samples changed only slightly from grade to grade, as the

population changed.

TABLE 1

Means and Standard Deviations of Kindergarten Predictors Samples for Grades 1-4

Variable	Grade 1 N=478		Grade 2 N=440		Grade 3 N=373		Grade 4 N=368	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MRT Total	55.99	17.43	56.69	16.50	57.31	16.70	57.35	16.55
MRT Subtests								
Wd. Meaning	9.63	3.01	9.65	2.91	9.70	2.95	9.75	2.93
Listening	9.22	2.54	9.27	2.48	9.42	2.41	9.38	2.43
Matching	7.52	3.66	7.65	3.68	7.86	3.71	7.78	3.70
Alphabet	9.51	4.50	9.61	4.33	9.57	4.34	9.67	4.34
Numbers	12.58	4.84	12.77	4.59	12.96	4.63	12.92	4.53
Copying	7.53	3.78	7.75	3.70	7.80	3.65	7.84	3.68
DAMT	3.06	1.07	3.10	1.01	3.10	1.01	3.14	1.02
BRS	3.51	.73	3.51	.72	3.56	.71	3.57	.72
TR	49.93	27.46	49.67	26.93	51.86	26.38	52.74	25.85

The criteria were the Metropolitan Achievement Test (MAT) in grade one and the Stanford Achievement Test (SAT) in grades two through four. An average of the reading subtests on each instrument was computed for each grade. The means and the standard deviations of the criterion variables are shown in Table 2.

TABLE 2

Means and Standard Deviations of Criterion Variables Grade Equivalent Scores

Grade	Mean	SD	Grade Norm
1	2.16	.66	1.8
2	2.98	.83	2.8
3	4.11	1.37	3.8
4	5.43	1.81	4.8

The means of the criteria were higher than the published test grade norms for each grade level examined, with a range from about two months higher, for second grade, to six months higher, for fourth grade. The school district population may be categorized as being normally distributed on the dependent variables.

The coefficients of correlation between the predictors and reading achievement are presented in Table 3.

TABLE 3

Correlations** between Predictors and Criteria

Variable	Grade 1	Grade 2	Grade 3	Grade 4
MRT Total	.63	.70	.74	.72
MRT Subtests				
Wd. Meaning	.44	.53	.61	.55
Listening	.31	.33	.37	.38
Matching	.50	.46	.51	.50
Alphabet	.63	.67	.68	.65
Numbers	.50	.58	.62	.59
Copying	.49	.51	.53	.53
DAMT	.39	.39	.39	.39
BRS	.42	.53	.49	.52
TR	.49	.53	.54	.52

** $p < .01$.

As would be expected, the correlation with achievement of the total MRT was higher than that of the subtests of this instrument. Indeed, the total MRT was the best predictor for all four grades, with the best prediction obtained for third grade, .74. The DAMT was the poorest predictor of the four for all the grades, with

a correlation of .39. The BRS and TR had some value, with correlations in the .40's and .50's.

Of the Metropolitan subtests, the Alphabet subtest was the best predictor and the Listening subtest was the poorest. After the first grade, none of the subtests had correlations of the magnitude of that for the total test.

The intercorrelations of the predictors are shown in Tables 4 and 5.

TABLE 4

Intercorrelations** of Four Predictors

Variable	DAMT	BRS	TR
	Grade 1 Sample		
MRT	.55	.66	.66
DAMT		.49	.46
BRS			.67
	Grade 2 Sample		
MRT	.49	.63	.63
DAMT		.44	.40
BRS			.64
	Grade 3 Sample		
MRT	.49	.64	.63
DAMT		.44	.38
BRS			.64
	Grade 4 Sample		
MRT	.50	.62	.59
DAMT		.47	.39
BRS			.63

** $p < .01$

TABLE 5

Intercorrelations ** of Nine Predictors

Variable	L	M	A	N	C	DAMT	BRS	TR
				Grade 1 Sample				
Word Meaning	.47	.52	.54	.64	.42	.39	.53	.50
Listening		.42	.31	.44	.31	.30	.37	.42
Matching			.54	.64	.56	.45	.51	.51
Alphabet				.64	.55	.42	.51	.53
Numbers					.59	.49	.58	.56
Copying						.47	.50	.53
DAMT							.49	.46
BRS								.67
				Grade 2 Sample				
Word Meaning	.41	.44	.51	.58	.35	.30	.47	.43
Listening		.38	.27	.39	.26	.25	.35	.37
Matching			.49	.63	.54	.40	.46	.47
Alphabet				.63	.53	.35	.49	.49
Numbers					.57	.43	.55	.51
Copying						.44	.48	.53
DAMT							.44	.40
BRS								.64
				Grade 3 Sample				
Word Meaning	.42	.45	.53	.61	.37	.32	.47	.45
Listening		.41	.27	.41	.28	.24	.37	.36
Matching			.51	.63	.57	.39	.47	.45
Alphabet				.65	.54	.38	.50	.53
Numbers					.56	.42	.57	.52
Copying						.46	.50	.55
DAMT							.44	.38
BRS								.64
				Grade 4 Sample				
Word Meaning	.42	.45	.52	.59	.37	.31	.46	.41
Listening		.41	.30	.39	.29	.26	.33	.32
Matching			.49	.63	.57	.39	.46	.40
Alphabet				.63	.54	.36	.48	.48
Numbers					.57	.43	.55	.48
Copying						.46	.51	.54
DAMT							.47	.39
BRS								.63

** $p < .01$

Inspection of these intercorrelations reveals a fairly high relationship between the MRT and the BRS and TR, with the magnitude of .60 and above. BRS and TR were also related, with a range of coefficients from .63 to .67. Thus these three instruments were measuring somewhat the same factors. Because the predictive ability of the MRT was greater than that of the other variables, as shown in Table 3, it may be considered to have measured some of what the two other predictors, BRS and TR, measured, and more. The DAMT showed little relationship to the other predictors, but it also had low predictive validity, limiting its value in multiple correlations.

Among the Subtests of the Metropolitan, the highest intercorrelations, generally in the .60's, were for Numbers with Word Meaning, Alphabet, and Matching. Correlations for Listening were generally lowest, with those for Copying somewhat higher. Listening was also a poor predictor for all grade levels.

Multiple correlations were computed to determine which variables best contributed to prediction, and to ascertain what percentage of the variance could be attributed to each predictor in combination with other predictors. These multiple correlations are reported in Table 6. It may be observed that MRT and TR were the chief contributors to the R, with the contributions of the DAMT and BRS negligible. The MRT contributed from about 40% to 55% of the variance. The variance

attributable to TR was much lower, from less than 1% to a maximum of 1.52%.

TABLE 6

Multiple Correlations** and Contributions of Four Variables

Item	Grade 1	Grade 2	Grade 3	Grade 4
R	.645	.714	.747	.728
% of variance due to:				
MRT	40.24	48.91	54.94	51.39
DAMT	.14	.06	.06	.05
BRS	.20	.47	.02	.18
TR	.99	1.52	.78	1.43
Total variance accounted for	41.54	50.93	55.78	53.02

** $p < .01$

When the Metropolitan subtests were used instead of the total MRT scores, the predictive contributions of the variables differed for the separate grade levels, unlike the relatively consistent contributions of MRT and TR in the four-variable correlation. The multiple correlations and contributions of the nine variables to the variance are reported in Table 7.

TABLE 7

Multiple Correlations** and Contributions of Nine Variables

Item	Grade 1	Grade 2	Grade 3	Grade 4
R	.681	.747	.772	.745
% of variance due to:				
MRT Subtests				
Word Meaning	.10	2.88	8.66	5.89
Listening	.05	.12	.27	.56
Matching	3.61	.01	.03	.08
Alphabet	39.88	45.35	46.33	42.82
Numbers	.02	.31	.56	.75
Copying	.54	.85	2.74	3.49
DAMT	.16	.21	.13	.13
BRS	.14	.59	.02	.25
TR	1.87	5.44	.86	1.59
Total variance accounted for	46.34	55.72	59.57	55.49

** $p < .01$.

In the first grade the Alphabet subtest was by far the best predictor, contributing almost 40% of the total 46% of variance accounted for. The Matching subtest and Teacher Ranking contributed more than 1% each. While these contributions were significant, they were hardly important. The six other variables contributed considerably less.

For the second grade the Alphabet subtest contribution was again the greatest (over 45%). The contribution of Teacher Ranking was greater for this grade than its contribution for first grade (about 5.5%), and Word Meaning, whose contribution was extremely small in first grade, increased substantially (from .10% to 2.88%). The

multiple prediction was better for second grade (.747) than for first (.681).

The multiple R for the third grade was the highest of all the grades, .772. Again the Alphabet subtest contributed the greatest proportion of the variance (46.33%). Word Meaning increased in importance (to 8.60%), as did Copying (2.74%), making these three the most important. TR diminished in its contribution to less than 1%.

The multiple R was somewhat lower for fourth grade, .728, with the Alphabet subtest continuing to contribute the greatest portion of the variance (close to 43%). For this grade, Word Meaning was also second (almost 6%), followed by Copying (3.5%), with TR increasing its contribution to more than 1% after having contributed very little in third grade. The other five variables contributed less than 1% each.

Constants and weights for multiple regression equations were obtained for each grade level so that predicted reading achievement scores could be computed for each child on the basis of the kindergarten measures. These are reported for the four variables in Table 8.

TABLE 8

R, Constants, and b Coefficients for Four-Variable Correlation

Item	Grade 1	Grade 2	Grade 3	Grade 4
R	.645**	.714**	.747**	.728**
Constants	9.398	7.718	6.789	6.072
b Coefficient				
MRT	.211**	.280**	.546**	.657**
DAMT	.283	.226	.387	.432
BRS	-.066	.105	-.389	.131
TR	.037**	.036*	.061*	.086*

** $p < .01$
* $p < .05$

The weights for the MRT were significant at all grade levels at the .01 level, and those for TR were significant at the .01 level for grade one and at the .05 level for the other grades.

For practical reasons it would seem advantageous to use regression equations including only the two most powerful predictors. Generally, addition of independent variables beyond the third contributes very little to efficiency, and in this instance there was very little added by DAMT and BRS. Indeed, there was no statistically significant difference between the four variable correlation coefficient and the two variable one on any of the four grade levels, as determined by an F-test, using the formula for the significance of difference between multiple R's:

$$F = \frac{(R^2_1 - R^2_2)(N - m - 1)}{(1 - R^2_1)(m_1 - m_2)}$$

where R_1 = multiple R with larger number of independent variables
R_2 = multiple R with one or more variables omitted
m_1 = larger number of independent variables
m_2 = smaller number of independent variables
$df_1 = (m_1 - m_2)$
$df_2 = (N - m_1 - 1)$

(Guilford, 1956, p. 403)

In fact, the exclusion of TR caused little loss of prediction when the R was compared with the Peason r. Even more practically, if the computation were to be done for many children, the MRT could be used alone. The constants and b coefficients for the two major predictors are reported in Table 9.

TABLE 9

Correlations, Constants, and b Coefficients for Two Major Predictors of Four

Item	Grade 1	Grade 2	Grade 3	Grade 4
r with MRT	.63**	.70**	.74**	.72**
R with 4 variables	.645**	.714**	.747**	.728**
R with 2 variables	.642**	.710**	.746**	.727**
Constant	8.387	10.199	6.538	9.297
b Coefficient				
MRT	.208**	.302**	.549*	.691**
TR	.032**	.049**	.059*	.103*

** $p < .01$.
* $p < .05$.

The constants and coefficients obtained when nine variables were employed are presented in Table 10.

TABLE 10

R, Constants, and b Coefficients
for Nine-Variable Correlations

Item	Grade 1	Grade 2	Grade 3	Grade 4
R	.681**	.747**	.772**	.745**
Constant	10.617	7.933	5.878	5.249
b Coefficient				
MRT Subtests				
Word Meaning	.082	.426**	.116**	.985**
Listening	.072	.136	.311	.549
Matching	.240**	- .030	.093	.187
Alphabet	.619**	.752**	1.076**	1.377**
Numbers	- .026	.131	.268	.331
Copying	.166*	.169	.418*	.498*
DAMT	.336	.436	.579	.714
BRS	- .054	.094	- .036	.143
TR	.036	.036**	.052*	.077*

** $p < .01$.
* $p < .05$.

For first grade the variables with significant weights were Alphabet, Matching, and Copying. Alphabet, Teacher Ranking, and Word Meaning had significant weights in second grade, and Alphabet, Word Meaning, Copying, and Teacher Ranking in third. For fourth grade significant weights were found for the same variables as for third grade.

The nine-variable correlation gave multiple correlations higher than those obtained using four variables, although it must be remembered that reliability of subtests is generally suspect (Cronbach, 1970, pp. 167-168). These multiple correlations are shown in Table 11.

TABLE 11

Comparison of R's Using Four and Nine Variables

Variables	Grade 1	Grade 2	Grade 3	Grade 4
Four	.645	.714	.747	.728
Nine	.681	.747	.772	.745

In both analyses, the best prediction was obtained for the third grade and the poorest for the first grade, although a coefficient in the .60's should not be considered a poor one. The kindergarten measures chosen seemed to predict best over a longer time period, rather than for the more immediate first grade. The use of Guilford's formula (1956, p. 403) showed that there was a statistically significant difference between these R's for all four grade levels in favor of the nine-variable regression equation.

Multiple Discriminant Analysis

The multiple regression study demonstrated which variables best predicted success in reading achievement for each of the grades tested. Since another purpose of the study was to determine those variables that best distinguished among reading achievement levels within each grade, a multiple discriminant analysis of the available data was included.

Before using discriminant analysis, it was first necessary to differentiate the students in each grade with respect to achievement, and then test, by use of Wilks'

Lambda, whether these achievement groups were actually separated on the predictor variables.

The division of the subjects into achievement groups is shown in Table 12.

TABLE 12

Scores of Achievement Groups

Grade	(Low) First Quartile	(Average) Middle 50%	(High) Fourth Quartile	Grade Norm
1	1.1 - 1.8	1.8 - 2.9	2.9 - 4.3	1.8
2	1.2 - 2.3	2.3 - 3.7	3.7 - 3.9	2.8
3	1.1 - 3.1	3.1 - 5.0	5.0 - 7.5	3.8
4	2.4 - 4.0	4.0 - 6.8	6.8 - 9.5	4.8

Table 13 reports the tests for equality of centroids, showing that the groups were in fact separated by the kindergarten variables. The degrees of freedom associated with the first and second roots were p(g-1) and (g-1)(N-g-p), respectively, where p=variates, g=groups, and N=subjects (Cooley & Lohnes, 1962, p. 61).

TABLE 13

Wilks' Lambda Test of Equality of Centroids

Grade	Λ	F-ratio**		df
		Four-Variable Analysis		
1	.667	26.523	8	944
2	.603	31.233	8	868
3	.694	18.411	8	734
4	.721	16.178	8	728
		Nine-Variable Analysis		
1	.626	13.712	18	934
2	.551	16.555	18	858
3	.517	15.700	18	724
4	.539	14.454	18	718

** $p < .01$.

Tables 14 through 17 show the group means and F-ratios for the significance of group differences on each of the variables.

TABLE 14

Group Means and F-ratios - Grade 1 Sample

Variable	Low Group Mean	Average Group Mean	High Group Mean	F-ratio**
MRT	42.563	55.461	70.358	111.468
MRT Subtests				
Word Meaning	8.008	9.582	11.317	42.440
Listening	8.235	9.301	10.042	16.376
Matching	5.294	7.343	10.092	66.023
Alphabet	6.067	9.464	13.062	101.283
Numbers	9.765	12.352	15.825	58.939
Copying	5.193	7.418	10.067	62.947
DAMT	2.555	3.063	3.558	29.474
BRS	31.067	35.151	39.033	41.323
TR	31.857	50.347	67.017	61.461

** $p < .01$.

TABLE 15

Group Means and F-ratios - Grade 2 Sample

Variable	Low Group Mean	Average Group Mean	High Group Mean	F-ratio**
MRT	41.491	57.345	70.336	132.468
MRT Subtests				
Word Meaning	7.636	9.691	11.518	61.333
Listening	8.173	9.373	10.109	18.120
Matching	5.455	7.609	9.854	47.266
Alphabet	5.546	9.918	12.991	130.099
Numbers	9.418	12.841	16.009	76.189
Copying	5.300	7.914	9.854	51.767
DAMT	2.546	3.159	3.536	30.488
BRS	29.864	35.809	38.882	53.936
TR	30.500	50.827	66.464	63.394

** $p < .01$.

TABLE 16

Group Means and F-ratios - Grade 3 Sample

Variable	Low Group Mean	Average Group Mean	High Group Mean	F-ratio**
MRT	47.946	59.759	73.893	8.069
MRT Subtests				
Word Meaning	8.419	9.481	12.290	15.666
Listening	8.355	9.454	10.484	19.023
Matching	6.677	7.433	10.785	15.725
Alphabet	5.602	9.824	13.215	106.097
Numbers	9.527	12.947	16.839	67.784
Copying	5.129	8.091	10.280	41.623
DAMT	2.516	3.193	3.548	29.358
BRS	30.892	35.995	40.161	41.939
TR	32.441	52.193	70.366	64.688

** $p < .01$.

TABLE 17

Group Means and F-ratios - Grade 4 Sample

Variable	Low Group Mean	Average Group Mean	High Group Mean	F-ratio**
MRT	44.086	56.136	79.022	31.153
MRT Subtests				
Word Meaning	7.957	9.603	11.839	53.025
Listening	8.118	9.419	10.548	26.288
Matching	6.151	7.310	10.409	41.650
Alphabet	6.097	9.614	13.409	102.720
Numbers	10.108	12.565	16.828	59.411
Copying	5.699	7.625	10.903	43.646
DAMT	2.699	3.125	3.645	22.385
BRS	30.774	35.935	40.462	44.963
TR	36.172	53.505	67.495	41.858

** $p < .01$.

Since the Wilks' lambda was significant, the next step was to apply the discriminant analysis. The results are shown in Tables 18 and 19. Chi square tests, computed from formulae derived by Rao (1952), were used to test the significance of the discriminant functions. Degrees of freedom were (p+g-2) and (p+g-4), respectively, for the two functions.

TABLE 18

Discriminant Analysis Using Four Variables

Grade	Trace	Root	% of Variance	χ^2	df
1	.499	1	99.38	190.947**	5
		2	.62	1.469	3
2	.653	1	98.66	217.061**	5
		2	1.34	3.794	3
3	.439	1	98.58	132.847**	5
		2	1.42	2.299	3
4	.381	1	95.66	113.931**	5
		2	4.34	6.011	3

** $p < .01$.

TABLE 19

Discriminant Analysis Using Nine Variables

Grade	Trace	Root	% of Variance	χ^2	df
1	.593	1	98.28	216.583**	10
		2	1.72	4.772	8
2	.801	1	97.80	251.160**	10
		2	2.20	7.596	8
3	.894	1	94.85	225.354**	10
		2	5.15	16.526*	8
4	.821	1	94.45	208.912**	10
		2	5.55	16.203*	8

** $p < .01$.
* $p < .05$.

These analyses could result in a maximum of two functions, as explained in the previous chapter. According to one authority:

> If a single discriminant function only were desired, the function associated with the first root was identical to what would have been obtained through the use of single discriminant analysis, that single function described by Fisher that best separates the groups by maximizing the ratio of the sum of squares between the groups to the sums of squares within the groups (Waldron, 1964, pp. 76-77).

The second discriminant function:

> ...maximizes the ratio of the residual among-groups sum-of-squares to the residual within-groups sum-of-squares after the effect of the first linear combination has been removed (Tatsuoka & Tiedeman, 1954, p. 410).

When four predictors were used, the first discriminant function alone was found to be significant. The groups were not separated on the second function at any grade level, demonstrating collinearity of the variables. In this analysis, the four kindergarten predictors, taken as a group, exhibited the same relationship with all three achievement groups. The Low group had the lowest centroid, the Average group a higher centroid, and the High group had the highest centroid on the first function, with no significant distances separating them on the second discriminant function. After the effect of this first linear combination, to which Tatsuoka and Tiedeman (1954) referred, there was no significant residual effect.

In the nine-variable analysis in which Metropolitan subtests were used, the first function was again significant at the .01 level for all four grades, and, in addition, the second function was significant at the .05 level for third and fourth grades.

Correlations of these functions with the variables are given in Table 20.

TABLE 20

Correlations of Discriminant Functions with Predictors

Variable	Grade I Functions		Grade 2 Functions		Grade 3 Functions		Grade 4 Functions	
	I	II	I	II	I	II	I	II
Four Predictors								
MRT	.982	-.164	.981	-.066	.368	.363	.723	.572
DAMT	.577	.261	.557	.327	.667	-.652	.637	.122
BRS	.668	.365	.704	.675	.782	.031	.856	-.239
TR	.786	.582	.758	.060	.925	.308	.830	-.330
Nine Predictors								
Wd.Meaning	.642	-.019	.705	.218	.378	.532	.715	-.145
Listening	.414	.395	.417	-.137	.449	.114	.533	.156
Matching	.767	-.339	.632	.375	.358	.666	.628	-.546
Alphabet	.901	.076	.921	-.093	.891	.071	.906	.046
Numbers	.733	-.326	.765	.291	.756	.363	.741	-.341
Copying	.754	-.144	.660	-.005	.633	-.008	.657	-.290
DAMT	.547	.128	.527	-.171	.543	-.202	.499	-.043
BRS	.634	.182	.667	-.401	.634	.063	.669	.189
TR	.746	.300	.715	.039	.750	.175	.648	.243

These coefficients indicate the strength of the relationship between each variable and each of the two functions, and ". . . may be interpreted in much the same way as factor loadings to describe the discriminant dimensions in terms of the names of the original variables"

(Veldman, 1967, p. 272).

The correlations of the functions were greatest with the variables which proved to be the chief predictors in the correlation analysis. Yet there was considerable overlapping among variables since it may be seen that the first function had a fairly high correlation with most of the variables. For the second function the correlations were generally low in terms of practical significance.

For the first two grades, the correlation of the first function with the total MRT was almost perfect. TR had the highest correlation with this function in grade three, with the MRT the lowest, .368. In fourth grade BRS and TR had the highest correlations, and the MRT correlation, .723, was substantial.

In the nine-variable analysis the correlation of Alphabet with the first, significant function was higher than that of any other variable, in the .90's on all grade levels except third where it was .891. For the first two grades, and fourth, all but two of the variables had a correlation of over .60 with the first function. For third grade there were three variables with a correlation below .60.

The second discriminant function was also significant, but at the .05 level, for third and fourth grades. For the third grade, this function had fairly good correlations, .666 and .532, with Matching and Word Meaning, respectively.

For the fourth grade its only efficient correlation was with Matching. The correlations of the variables with the second function, it must be noted, were not of the magnitude of the correlations of these same variables with the first function.

For the two grades where both functions were significant, the relationships are shown in Figure 1 and 2.

L Low Group Centroid

A Average Group Centroid

H High Group Centroid

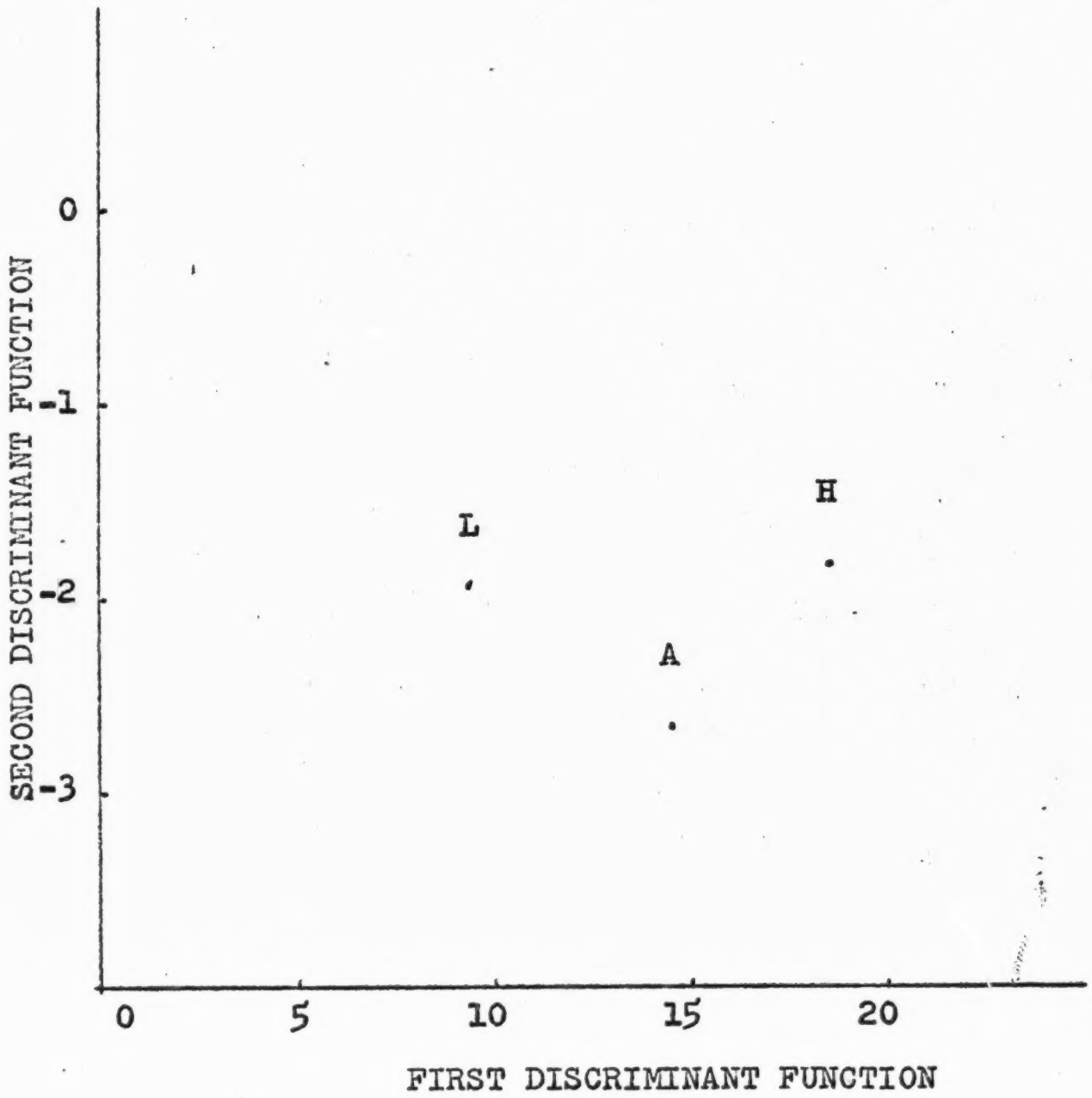

Fig. 1. Centroids of discriminant functions when both were significant in nine-variable analysis for Grade 3.

L Low Group Centroid

A Average Group Centroid

H High Group Centroid

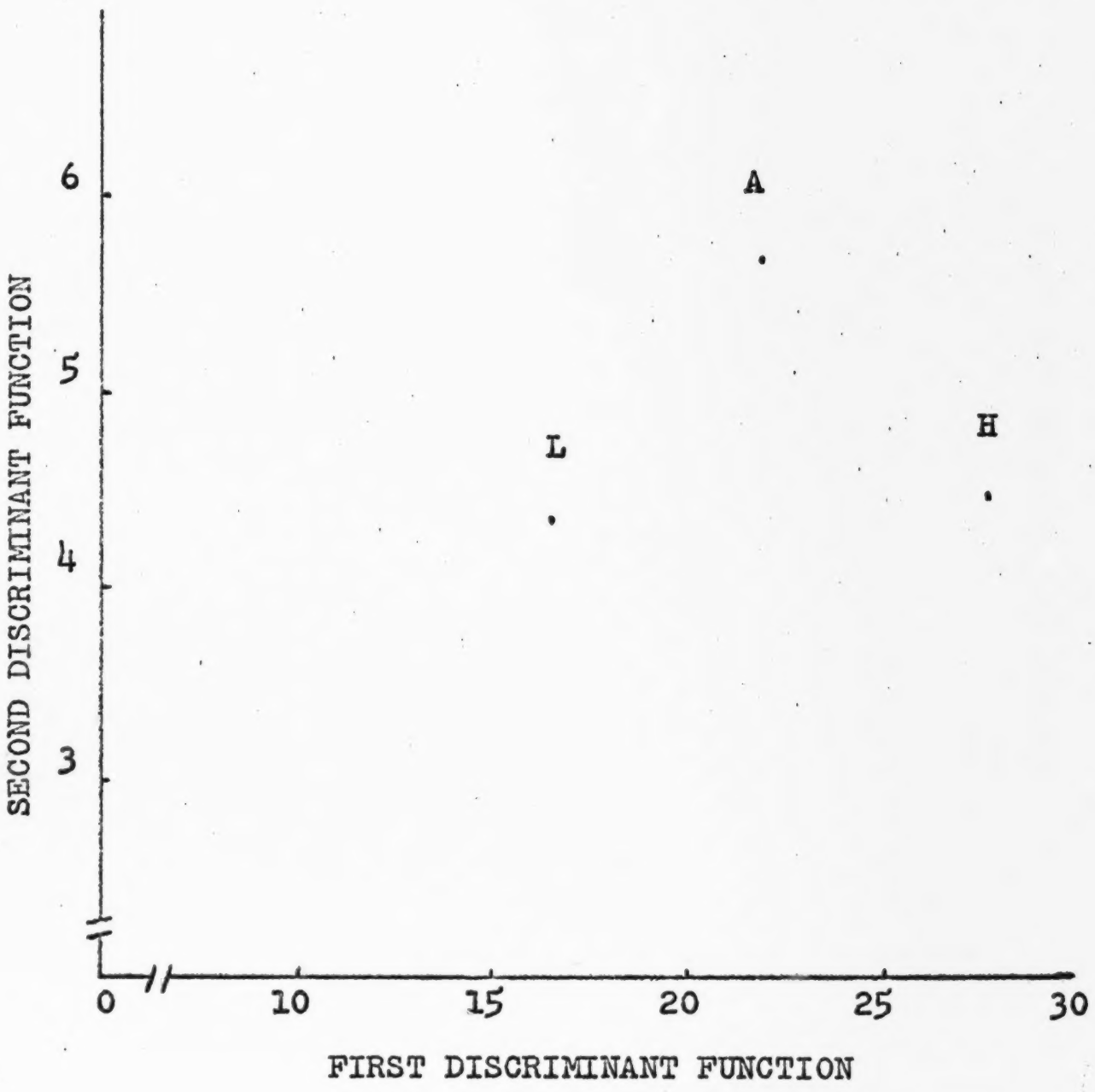

Fig. 2. Centroids of discriminant functions when both were significant in the nine-variable analysis for Grade 4.

It may be observed that in both these grades, the Average group was distinctly separated from the Low and High groups on the second discriminant function while all three were separated on the first function. There was little separation between Low and High on the second function for either grade.

Scaled vectors were computed to show the relative contributions of the variables to group discrimination. These are shown in Table 21. The negative sign of some vectors does not influence the importance of the vectors; their contributions to discrimination are to be considered in terms of their absolute value only.

TABLE 21

Scaled Vectors

Variable	Grade 3		Grade 4	
	Function I	Function II	Function I	Function II
MRT Subtests				
Word Meaning	.145	.105	1.113	- .659
Listening	.673	- .571	1.007	1.166
Matching	- .706	1.148	.229	-2.316
Alphabet	2.915	-1.132	3.369	1.383
Numbers	.802	1.434	.137	-1.300
Copying	.335	- .441	.613	1.078
DAMT	.622	- .822	.211	.054
BRS	- .021	- .048	.056	.166
TR	1.276	.690	.514	1.069

The three most important variables for the first function for third grade, when ranked in order of importance, were Alphabet, Teacher Ranking, and Numbers. The second function on this grade level depended on Numbers, Matching, and Alphabet as its three most

important variables. Alphabet and Numbers made important contributions to both functions for third grade.

For fourth grade Alphabet, Word Meaning, and Listening contributed most to the first function, with Matching most important to the second function, followed by Alphabet, Numbers, Letters, Copying, and Teacher Ranking, which were all relatively close in their contributions. Alphabet and Listening were important to both functions for this grade level.

The mean scores for all the variables mentioned as important for these two grades were low for the Low group, higher for the Average group, and highest for the High group.

An examination of the means and F-ratios would not have given the information as to which predictors best separated these groups, since all the F-ratios were significant.

Thus discriminant function analysis gave information which could not have been obtained by the use of analysis of variance or multiple correlation alone.

This investigation showed that prediction of later reading achievement was in fact possible from kindergarten measures.

CHAPTER V

Summary, Conclusions, and Recommendations

Summary

This investigation sought to discover whether it was possible to predict reading achievement in the first four grades by means of readily available kindergarten measures. The multivariate analysis was used to answer four general questions:

1. What is the extent to which kindergarten measures can predict reading achievement in grades one, two, three, and four?

2. Are these measures equally effective in predicting achievement on the four grade levels, or are different measures better predictors of reading achievement in different grades?

3. Do correlations among the measures indicate that duplication of measurement exists, or are the measures sufficiently independent to warrant their inclusion in such a battery?

4. Which composite of kindergarten measures best discriminates the high, average, and low achieving readers at each grade level?

The subjects of the study were 553 children in the Ithaca City School District public schools kindergarten in 1964 for whom there were one or more reading achievement test scores in grades one through four. There were 478 subjects in the first grade sample, 440 in

the second, 373 in the third, and 368 in the fourth grade sample.

The four predictive measures used were those routinely available to the kindergarten teachers, and did not require special training or abilities to administer or score. They were the Metropolitan Readiness Test (MRT) with its six subtests of Word Meaning, Listening, Matching, Alphabet, Numbers, and Copying; the Draw-a-Man Test (DAMT), a nonverbal intelligence test used with the MRT; the district's behavior rating scale (BRS), and a teacher ranking of a child's readiness within his class (TR).

Two techniques formed the basis for analysis of the data. A multiple correlation and regression analysis produced correlations of the individual predictors with the criteria, average reading achievement as measured by the Metropolitan Achievement Test in first grade and by the Stanford Achievement Test in the next three grades. Independent variables were used in two ways -- once with the MRT total score as one predictor along with the DAMT, BRS, and TR, and once with the six MRT subtests used as independent variables along with the DAMT, BRS, and TR. Intercorrelations of the predictors were also obtained. Multiple correlations indicated combinations of variables giving the best prediction, and regression coefficients and constants gave information for predicting individual achievement scores.

Discriminant analysis was used to determine which kindergarten variables predicted membership in high, average, and low achieving groups later in school.

The results of the study revealed that when four independent variables were used, the MRT was the best single predictor, with correlation coefficients ranging from .63 for grade one achievement to .74 for third grade. In comparison, the DAMT yielded a correlation of .39 for all four grades, the BRS ranged from .42 to .53, and TR from .49 to .54. These three predictors showed a substantial relationship to the MRT and to each other, but the MRT was a better predictor and it was assumed that while there was an overlap of measurement, it was the MRT that was measuring the most and could therefore be used as the single predictor.

Where the six MRT subtests were used, the Alphabet subtest had the highest correlation with the criteria on all grade levels, from .63 for first grade to .68 for third. For first grade this subtest had the same predictive power as the total test score, .63, but for later grades the total test increased in superiority over this subtest. In second, third, and fourth grades its correlation coefficients were .67, .68, and .65, respectively. The Numbers subtest, which has been cited as the best predictor of the subtests, was generally second as a predictor. The Listening subtest, with its correlations in the .30's, was the poorest, even lower than the DAMT.

In the multiple correlations using four variables, R's ranged from .643, for grade one, to .747 for grade three. (All correlations appeared to be lowest for first grade and highest for third, when the reading skill has generally been established.) On all grade levels the MRT accounted for 40% to almost 55% of the variance, with the other predictors accounting for little more than 1.5%.

Using nine variables, significantly higher multiple correlations were found, from .681 to .772. The Alphabet subtest contributed from about 40% to 45% of the variance, with various variables contributing up to 8.66%. Other variables contributed significantly to the multiple correlation, but in very small measure.

Thus it may be seen that the MRT stood alone as the best predictor for all grades, and its Alphabet subtest contributed most when MRT scores were analyzed.

Discriminant analysis with four variables revealed only one significant function with an extremely high correlation with the MRT in the first two grades. In subsequent grades other variables assumed more importance within this function. In the nine variable analysis the first function, significant at the .01 level for all grades, had a high relationship with the Alphabet subtest. In third and fourth grades the second function was significant at the .05 level. The average group in third grade was separated from the extreme group on this function, to which Numbers, Matching, and Alphabet contributed most. The same

kind of separation occurred in fourth grade where four subtests, Alphabet, Numbers, Listening, Copying, and TR contributed heavily to the second function. Alphabet was important to both functions in both grades.

Again it must be noted that the lack of reliability of subtests generally, when compared with a total test score, might caution against absolute reliance on the nine-variable discriminant analysis. The four-variable analyses demonstrated in a forthright way the predictive validity of the MRT.

Conclusions

The results of this investigation would seem to justify the following conclusions:

1. Kindergarten measures proved to be effective predictors of reading achievement in grades one, two, three, and four.

2. The best prediction of reading achievement was obtained for grade three, and the lowest, although good, was for first grade.

3. The predictor variables were correlated with each other so that duplication of measurement probably existed. Consequently it would not be necessary to include all the variables in a predictive battery.

4. The MRT was the best single predictor of later reading achievement, showing substantial correlation with the criteria on all four grade levels.

5. Multiple correlations using two or four variables

increased predictive ability only slightly over use of the MRT as the sole predictor.

6. Multiple correlations using the MRT subtests and the three other independent variables produced significantly higher, but less reliable, correlations, with the Alphabet subtest contributing the greatest proportion of the variance.

7. The poorest predictor, with a correlation of .39 for all grade levels, was the DAMT, a nonverbal test of intelligence.

8. Multiple discriminant analysis showed that the centroids in the four variable analysis were collinear. The groups were separated on one function only, which was generally related to all four variables.

9. The nine-variable multiple discriminant analysis showed two significant functions, one at .01 and the other at .05. The second function did not have correlations with the variables as high as those of the first function. At least half of the variables had substantial correlations with the first function.

Recommendations

1. Further study could determine whether a verbal test of intelligence, as opposed to the DAMT, would contribute more to multiple prediction or would merely duplicate some of the verbal factors of the MRT.

2. The multiple correlations found in this study have contributed at most 59.97% of the variance. Further

studies are needed to identify the variables that determine the composition of the other 40% of the variance.

3. The finding that the Alphabet subtest was a strong predictor and contributor to both functions on the two grade levels where two functions were significant deserves further investigation to determine what this test actually measures.

BIBLIOGRAPHY

Alshan, L. M. Reading readiness and reading achievement. In J. A. Figuerel (Ed.), Reading and inquiry. Newark, Delaware: International Reading Association, 1965.

Anderson, H. E. Regression, discriminant analysis, and a standard notation for basic statistics. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966. Pp. 153-173.

Anderson, I. H. The Metropolitan Readiness Test. In O. K. Buros (Ed.), Third mental measurements yearbook. New Brunswick, New Jersey: Rutgers University Press, 1949.

Andras, M. C. Growth curves and factor pattern changes in first grade reading readiness, reading achievement and its prediction. Dissertation Abstracts, 1965, 26, 4488-4489.

Ausubel, D. P. Viewpoints from related disciplines: human growth and development. Teachers College Record, 1959, 60, 245-254.

Bagford, J. Reading readiness scores and success in reading. The Reading Teacher, 1968, 28, 324-328.

Banham, K. M. Maturity level for reading-readiness. Educational and Psychological Measurement, 1958, 18, 371-375.

Barrett, T. C. The relationship between measures of pre-reading visual discrimination and first-grade reading achievement: a review of the literature. Reading Research Quarterly, 1965, 1, 51-75.

Benger, K. A study of the relationships between perception, personality, intelligence, and grade one reading achievement. Paper presented at the meeting of the International Reading Association, Seattle, 1967.

Blair, G. M. & Jones, R. S. Readiness. In C. W. Harris (Ed.), Encyclopedia of educational research. New York: Macmillan, 1960. Pp. 1081-1085.

Bobbe, C., Campbell, W., Lamberti, E., & Sheppard, C. A correlation analysis in testing. Education, 1963, 83, 375-378.

Bock, R. D. Contributions of multivariate experimental designs to educational research. In R. B. Cattel (Ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966. Pp. 820-840.

Bond, G. L. The coordinated phases of the reading study. Paper presented at the meeting of the International Reading Association, Dallas, May, 1966.

Bond, G. L. & Bond, E. Teaching the child to read. New York: Macmillan, 1945.

Bremer, N. Do readiness tests predict success in reading? Elementary School Journal, 1959, 59, 222-224.

Brill, M. The reliability of the Goodenough Draw-a-Man Test and the validity and reliability of an abbreviated scoring method. Journal of Educational Psychology, 1935, 26, 701-708.

Bruner, J. S. The process of education. Cambridge: Harvard University Press, 1960.

Bryan, J. G. A method for the exact determination of the characteristic equation and latent vectors of a metrix with applications to the discriminant function for more than two groups. Unpublished thesis, Harvard University, 1950.

Bryan, Q. R. Relative importance of intelligence and visual perception in predicting reading achievement. California Journal of Educational Research, 1964, 15, 44-48.

Carr, J. W., Jr., & Michaels, M. N. Reading readiness tests and grouping of first-grade entrants. Elementary English Review, 1941, 18, 133-138.

Castner, B. M. Prediction of reading disability prior to first grade entrance. American Journal of Orthopsychiatry, 1935, 5, 375-387.

Chall, J. Learning to read: the great debate. New York: McGraw-Hill, 1967.

Charry, L. B. The relationship between pre-reading and first grade reading performances and subsequent achievement in reading and other specified areas. (Doctoral dissertation, Temple University) Ann Arbor, Michigan: University Microfilms, 1967, No. 67-11418.

Cooley, W. W. & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: John Wiley, 1962.

Cooley, W. W. & Lohnes, P. R. Project TALENT. Predicting development of young adults. Palo Alto: American Institutes for Research and School of Education, University of Pittsburgh, 1968.

Cronhach, L. J. Essentials of psychological testing. (3rd ed.) New York: Harper & Row, 1970.

Crowley, F. J. & Cohen, M. Statistics for examination review. New York: Collier Books, 1967.

Danford, B. H. Some correlates of two brief intelligence tests used by pediatricians. Dissertation Abstracts, 1965, 26, 1772-1773.

Dean, C. D. Predicting first-grade reading achievement. Elementary School Journal, 1939, 39, 609-616.

Dean, E. K. Significant factors associated with reading achievement in the primary grades -- a longitudinal study. New York: The American Press, 1965.

de Hirsch, K., Jansky, J. J., & Langford, W. S. Predicting reading failure. New York: Harper & Row, 1966.

Deputy, E. C. Predicting first grade reading achievement. Teachers College Contributions to Education, No. 426. New York: Teachers College, Columbia University, 1930.

Dobson, J. C. & Hopkins, K. D. The reliability and predictive validity of the Lee-Clark Reading Readiness Test. Journal of Developmental Reading, 1963, 6, 278-281.

Durkin, D. A case-study approach toward an identification of factors associated with success and failure in learning to read. California Journal of Educational Research, 1960, 11, 26-33.

Durkin, D. A fifth year report on the achievement of early readers. Elementary School Journal, 1964, 65, 76-80.

Durkin, D. Children who read early. New York: Teachers College Press, Columbia University, 1966.

Durkin, D. When should children begin to read? In H. M. Robinson (Ed.), Innovation and change in reading instruction. The sixty-seventh yearbook of the National Society for the Study of Education, Part II, Chicago: N. S. S. E., 1968. Pp. 30-71.

Dykstra, R. The use of reading readiness tests for prediction and diagnosis: a critique. In T. C. Barrett (Ed.), The evaluation of children's reading achievement. Newark, Delaware: International Reading Association, 1967. Pp. 35-51.

Easley, G. T. The Draw-a-Man Test as an index of reading readiness. Dissertation Abstracts, 1964, 25, 2881.

Edwards, A. J. & Kirby, M. E. Predictive efficiency of intelligence test scores; intelligence quotients obtained in grade one and achievement test scores obtained in grade three. Educational and Psychological Measurement, 1964, 24, 941-946.

Fransella, F. Multiple regression equations for predicting reading age from chronological age and WISC verbal IQ. British Journal of Educational Psychology, 1965, 35, 86-89.

Gates, A. I. The necessary mental age for beginning reading. Elementary School Journal, 1937, 37, 497-508.

Gates, A. I. A further evaluation of reading-readiness tests. Elementary School Journal, 1940, 40, 577-591.

Gates, A. I. & Bond, G. L. Reading readiness -- a study of factors determining success and failure in beginning reading. Teachers College Record, 1936, 37, 679-685.

Gates, A. I., Bond, G. L., & Russell, D. H. Methods of determining reading readiness. New York: Bureau of Publications, Teachers College, Columbia University, 1939.

Gavel, S. June reading achievements of first-grade children. Journal of Education, 1958, 140, 37-48.

Gjessing, H. J. The concept of reading readiness in Norway. In M. D. Jenkinson (Ed.), Reading instruction: an international forum. Proceedings of the First World Congress on Reading, Paris, France, 1966. Newark, Delaware: International Reading Association, 1967. Pp. 70-79.

Griffith, F. The Metropolitan Readiness Test. Teachers College Journal, 1949, 21, 15-16.

Guilford, J. P. Fundamental statistics in psychology and and education. New York: McGraw-Hill, 1956.

Hampleman, R. S. A study of the comparative reading achievement of early and late starters. In A. J. Harris (Ed.), Readings on reading instruction. New York: McKay, 1963. Pp. 61-65.

Harootunian, B. Intellectual abilities and reading achievement. Elementary School Journal, 1966, 66, 386-392.

Harris, A. J. Psychological bases of reading in the United States. In M. D. Jenkinson (Ed.), Reading instruction: an international forum. Proceedings of the First World Congress on Reading, Paris, France, 1966. Newark, Delaware: International Reading Association, 1967. Pp. 336-349.

Harris, A. J. Diagnosis and remedial instruction. In H. M. Robinson (Ed.), Innovation and change in reading instruction. The sixty-seventh yearbook of the National Society for the Study of Education, Part II. Chicago: N. S. S. E., 1968. Pp. 159-194.

Harris, D. B. Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace, & World, 1963.

Henderson, E. H. & Long, B. H. Correlations of reading readiness among children of varying background. The Reading Teacher, 1968, 22, 40-44.

Henig, M. S. Predictive value of a reading readiness test and of teachers' forecasts. Elementary School Journal, 1949, 50, 40-46.

Hildreth, G. & Griffiths, N. Metropolitan Readiness Tests. New York: World Book, 1939.

Hildreth, G. H., Griffiths, N. L., & McGauvran, M. E. Manual of directions, Metropolitan Readiness Tests. New York: Harcourt, Brace, & World, 1965.

Hillerich, R. L. An interpretation of research in reading readiness. Elementary English, 1966, 43, 359-364, 372.

Hobson, J. R. The Metropolitan Achievement Tests. In O. K. Buros (Ed.), The fourth mental measurements yearbook. Highland Park, New Jersey: The Gryphon Press, 1953.

Hopkins, K. D. & Sitkei, E. G. Predicting grade one reading performance: intelligence versus reading readiness tests. Paper presented at the meeting of the American Educational Research Association, New York, February, 1967.

Inglis, W. B. The early stages of reading: a review of recent investigations. In The Scottish Council for Research in Education (Ed.), Studies in reading, Vol. 1. London: University of London Press, 1948. Pp. 1-92.

Jackson, J. A survey of psychological, social, and environmental differences between advanced and retarded readers. Journal of Genetic Psychology, 1944, 65, 113-121.

Johnson, M. S. Factors related to disability in reading. Journal of Experimental Education, 1957, 26, 1-26.

Karlin, R. The prediction of reading success and reading readiness tests. Elementary English, 1957, 34, 320-322.

Kelley, T. L., Madden, R., Gardner, E. F., & Rudman, H. C. Stanford Achievement Test, directions for administering. New York: Harcourt, Brace, & World, 1965.

Keogh, B. K. Form copying tests for prediction of first grade reading. In M. P. Douglass (Ed.), Claremont Reading Conference twenty-seventh yearbook, 1963. Pp. 141-144.

Kerfoot, J. F. The relationship of selected auditory and visual reading readiness measures to first grade reading achievement and second grade reading and spelling achievement. Dissertation Abstracts, 1964, 25, 1747-1748.

Kermoian, S. G. Teacher appraisal of first grade readiness. Elementary English, 1962, 39, 196-201.

Kingston, A. J., Jr. The relationship of first-grade readiness to third- and fourth-grade achievement. Journal of Educational Research, 1962, 56, 61-67.

Koppitz, E., Sullivan, J., Blyth, D., & Shelton, J. Prediction of first-grade school achievement with the Bender Gestalt Test and human figure drawings. Journal of Clinical Psychology, 1959, 15, 164-168.

Kottmeyer, W. Handbook for remedial reading. St. Louis: Webster Publishing Company, 1947.(a)

Kottmeyer, W. Readiness for reading. Elementary English, 1947, 24, 355-366(b)

Kottmeyer, W. Readiness for reading. In W. B. Barbe (Ed.), Teaching reading: selected materials. New York: Oxford University Press, 1965, 81-87.

League of Women Voters of Tompkins County. Know your schools: Ithaca City School District. Ithaca: LWV of Tompkins County, 1967.

Lee, J. M., Clark, W. W., & Lee, D. M. Measuring reading readiness. Elementary School Journal, 1934, 34, 656-666.

Leton, D. A. A factor analysis of readiness tests. Perceptual and Motor Skills, 1963, 16, 915-919.

Matick, W. E. Predicting success in the first grade. Elementary School Journal, 1963, 63, 273-276.

Mayans, A. E. Early differential prediction of first grade reading achievement among three culturally different kindergarten groups. (Doctoral dissertation, University of Cincinnati) Ann Arbor, Michigan: University Microfilms, 1966. No. 67-1998.

McCarthy, D. A study of the reliability of the Goodenough Drawing Test of Intelligence. Psychological Bulletin, 1937, 34, 459-460.

McHugh, G. Changes in Goodenough IQ. at the public school kindergarten level. Journal of Educational Psychology, 1945, 36, 17-30.

McKim, M. G. The Metropolitan Achievement Tests. In O. K. Buros (Ed.), The fourth mental measurements yearbook. Highland Park, New Jersey: The Gryphon Press, 1953.

Medinnus, G. R. The development of a first grade adjustment scale. Journal of Experimental Education, 1961, 30, 243-248.

Meyers, C. E., Attwell, A. A., & Orphet, R. E. Prediction of fifth grade reading achievement from kindergarten test and rating data. Educational and Psychological Measurement, 1968, 28, 457-463.

Mitchell, B. C. The Metropolitan Readiness Tests as predictors of first grade achievement. Educational and Psychological Measurement, 1962, 22, 765-772.

Mitchell, B. C. Predictive validity of the Metropolitan Readiness Tests and the Murphy-Durrell Reading Readiness analysis for white and Negro pupils. Educational and Psychological Measurement, 1967, 27, 1047-1054.

Monroe, M. Reading aptitude tests for the prediction of success and failure in beginning reading. Education, 1935, 56, 7-14.

Morean, M. Long term prediction of reading success. California Journal of Educational Research, 1950, 1, 173-176.

Morphett, M. V. & Washburne, C. When should children begin to read? Elementary School Journal, 1931, 31, 496-503.

Mullis, J. C. The prediction of fifth grade achievement as measured by teacher grades and achievement test scores using first grade indices of prediction. Dissertation Abstracts, 1966, 26, 6515.

Neville, D. A comparison of the WISC patterns of male retarded and non-retarded readers. Journal of Educational Research, 1961, 54, 195-197.

Nicholson, A. Background abilities related to reading success in first grade. Journal of Education, 1958, 140, 7-24.

Olson, A. V. Growth in word perception abilities as it related to success in beginning reading. Journal of Education, 1958, 140, 35-36.

Orear, M. L. Social maturity and first grade achievement. California Journal of Educational Research, 1951, 2, 84-88.

Osburn, W. J. The Metropolitan Readiness Tests. In O. K. Buros (Ed.), 1940 mental measurements yearbook. Highland Park, New Jersey: The MMY, 1940.

Panther, E. E. Prediction of first-grade reading achievement. Elementary School Journal, 1967, 68, 44-48.

Petty, M. C. An experimental study of certain factors influencing reading readiness. Journal of Educational Psychology, 1939, 30, 215-231.

Pharis, W. L., Jr. The development of a rating scale to predict achievement in kindergarten. Dissertation Abstracts, 1967, 27, 3307A.

Powell, M. & Parsley, K. M., Jr. The relationships between first grade reading readiness and second grade reading achievement. Journal of Educational Research, 1961, 54, 229-233.

Ransom, G. A. First grade screening test. Journal of Educational Measurement, 1969, 6, 36-37.

Rao, C. R. Advanced statistical methods in biometric research. New York: John Wiley & Sons, 1952.

Robinson, H. A. Reliability of measures related to reading success of average, disadvantaged, and advantaged kindergarten children. The Reading Teacher, 1966, 20, 203-209.

Robinson, H. A. & Hanson, E. Reliability of measures of reading achievement. The Reading Teacher, 1968, 21, 307-313, 323.

Robinson, H. M. Factors affecting success in reading. Elementary School Journal, 1955, 55, 263-269.

Robinson, H. M. The Stanford Achievement Tests. In O. K. Buros (Ed.), The fifth mental measurements yearbook. Highland Park, New Jersey: The Gryphon Press, 1959.

Robinson, H. M. Weintraub, S., & Hostetter, C. Summary of investigations relating to reading, July 1, 1963 to June 30, 1964. The Reading Teacher, 1965, 18, 383-399.

Roslow, S. Reading readiness and reading achievement in first grade. Journal of Experimental Education, 1940, 9, 154-159.

Santoro, R. M. The relationship of reading achievement to specific measures of visual perception, visual-motor perception and intelligence. Unpublished doctoral dissertation, Fordham University, 1967.

Shea, C. A. Visual discrimination of words and reading readiness. The Reading Teacher, 1968, 21, 361-367.

Shipp, D. E. & Loudon, M. L. The Draw-a-Man Test and achievement in the first grade. Journal of Educational Research, 1964, 57, 518-521.

Silberberg, N., Iversen, I., & Silberberg, M. Predicting end of first grade Developmental Reading Test scores from Gates Reading Readiness Test scores administered in kindergarten. Paper presented at the meeting of the American Psychological Association, Washington, 1967.

Silberberg, N., Iversen, I., & Silberberg, M. The predictive efficiency of the Gates Reading Readiness Tests. Elementary School Journal, 1968, 68, 213-218.

Silvaroli, N. J. Intellectual emotional factors as predictors of children's success in first grade reading. Dissertation Abstracts, 1964, 24, 5098.

Slobodzian, E. The relationship between certain readiness measures and reading achievement at level one. (Doctoral dissertation, Temple University) Ann Arbor, Michigan: University Microfilms, 1963, No. 68-14151.

Smith, F. O. What the Goodenough Intelligence Test measures. Psychological Bulletin, 1937, 34, 760-761.

Sprigle, H. A., & Lanier, J. Validation and standardization of a school readiness screening test. Journal of Pediatrics, 1967, 70, 602-607.

Stewart, N. The Goodenough Intelligence Test. In O. K. Buros (Ed.), The fourth mental measurements yearbook. Highland Park, New Jersey: The Gryphon Press, 1953.

Strang, R., McCullough, C. M., & Traxler, A. E. The improvement of reading. New York: McGraw-Hill, 1961.

Tatsuoka, M. M. & Tiedeman, D. V. Discriminant analysis. Review of Educational Research, 1954, 24, 402-420.

Thackray, D. V. The relationship between reading readiness and reading progress. British Journal of Educational Psychology, 1965, 35, 252-254.

Tyler, T. F., Jr. The relationship of teacher judgments and environmental data at the kindergarten and grade one levels to reading achievement in grade four. Dissertation Abstracts, 1968, 28, 2151A.

Veldman, D. J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart, & Winston, 1967.

Vilscek, E. C. An analysis of the effects of mental age levels and socioeconomic status on reading achievement in first grade. Dissertation Abstracts, 1965, 26, 913-914.

Waldron, Brother Cormac. Differential prediction of achievement in broad curricular areas in an academic high school. (Doctoral dissertation, Fordham University) Ann Arbor, Michigan: University Microfilms, 1964. No. 64-8595.

Wartenberg, H. The relationship between success in beginning reading and various predictive measures. (Doctoral dissertation, Temple University) Ann Arbor, Michigan: University Microfilms, 1967. No. 67-11442.

Weaver, A. S. The prediction of first grade reading achievement in culturally disadvantaged children. Disseration Asbstracts, 1968, 28, 3789A.

Weintraub, S. What research says to the reading teacher. The Reading Teacher, 1967, 20, 551-558.

Whipple, G. The concept of reading readiness in the United States of America. In M. D. Jenkinson (Ed.), Reading instruction: an international forum. Proceedings of the First Word Congress on Reading. Newark, Delaware: International Reading Association, 1967. Pp. 80-90.

Wilking, S. V. Personality maladjustment as a causative factor in reading disability. Elementary School Journal, 1941, 42, 268-270.

Wilson, F. T. & Burke, A. Reading readiness in a progressive school. Teachers College Record, 1937, 38, 565-580.

Wilson, F. T. & Flemming, C. W. Correlations of reading progress with other abilities and traits in grade one. Journal of Genetic Psychology, 1938, 53, 33-52.

Wilson, F. T., Flemming, C. W., Burke, A., & Garrison, C. G. Reading progress in kindergarten and primary grades. Elementary School Journal, 1938, 38, 442-449.

Witty, P. & Kopel, D. Reading and the educative process. Boston: Ginn & Company, 1939.

Yule, W., Lockyer, L., & Noone, A. The reliability and validity of the Goodenough-Harris Drawing Test. British Journal of Educational Psychology, 1967, 37, 110-111.

Zaruba, E. A. Objective and subjective evaluation at grade one. The Reading Teacher, 1968, 22, 50-54.

Zieky, M. J. & Page, E. B. Review of de Hirsch, Jansky, & Langford, Predicting reading failure. Harvard Educational Review, 1968, 38, 365-366.

APPENDIX

Department of Psychological Services
Ithaca City School District
Ithaca, New York

BEHAVIOR RATING SCALE

Directions for Rating: Please read carefully the descriptive phrases which appear at either end of each trait scale. These phrases define the traits and are to be considered as anchor points. As will be noted, anchor point (1) describes the negative or undesirable aspect of a trait while the opposite is true of anchor point (5). Each trait should be thought of, then, as existing on a continuum from negative to positive. The center of each scale (3), indicates the position in which the "average" child would be placed with regard to a trait.

In making your decision concerning the rating to be given, freely compare the child with others in the group on a particular dimension. Consider only one trait at a time. In this way a generally favorable or unfavorable impression which you may hold of a child will not influence your objective rating of him on each scale. Although occasionally you may feel that the child is best described by a point lying between two numbers, please select one of the numbers.

After selecting one number as a rating, indicate the number on the separate blank provided opposite the appropriate item number. <u>Please do not write on the rating scale itself</u>.

Poor		Average		Excellent
(1)	(2)	(3)	(4)	(5)

I. <u>Motor and Speech Behavior</u>

1. (1) Poor physical coordination in large muscle activities, awkward ... (5) Excellent physical coordination; agile

2. (1) Poor control in fine muscle activities, drawing, coloring ... (5) Excellent muscular control, e.g. holding crayon, chalk

3. (1) Indistinct, infantile speech ... (5) Mature speech patterns, good enunciation

(1)	(2)	(3)	(4)	(5)
II. Social Behavior				
4.	(1) Doesn't get along well with other children, quarrels, fights, bullies, has few playmates	...	(5) Gets along well with other children, has many playmates, is well liked	
5.	(1) Inconsiderate; self-centered	...	(5) Considerate, thinks of others, respects abilities and opinions of others	
6.	(1) Unwilling to share, hasn't learned to share	...	(5) Always willing to share	
7.	(1) Doesn't know how or is unable to play in a large group	...	(5) Is able to play in a large group	
8.	(1) Defensiveness-every little hit, demands a full battle	...	(5) Doesn't have to fight back if pushed or hit, able to accept taunts	
9.	(1) Shy, bashful, timid initiates few contacts with other children	...	(5) Outgoing, initiates many contacts with other children	
10.	(1) Socially dependent; disturbs other children; handles, touches them, pokes them	...	(5) Socially independent and self-sufficient; doesn't touch or poke other children	
III. Emotional Behavior				
11.	(1) Unhappy, gloomy disposition, whiny, pouts	...	(5) Happy, cheerful	
12.	(1) Inflexible - gets upset over a change in the pattern of activities	...	(5) Flexible - if doesn't get own way can modify behavior, doesn't get upset over a change in routine	

(1)	(2)	(3)	(4)	(5)
13.	(1) Displays acting-out aggressiveness toward others	...	(5) Manages to keep angry feelings within acceptable limits	
14.	(1) Cries easily; becomes easily upset; lacks self-control; babyish	...	(5) Stable; possesses self-control	
15.	(1) Tense; nervous, lacks self-confidence	...	(5) Relaxed, self-confident	
16.	(1) Initial fear of school; took a long time to adjust to the school situation	...	(5) Adjusted readily to school; apprehension regarding school	

IV. Intellectual Abilities and Behavior

17.	(1) Immature language development, unable to express self adequately; poor vocabulary	...	(5) Mature language development; able to talk in simple accurate sentences, large vocabulary for age level	
18.	(1) Few original ideas; unimaginative	...	(5) Very creative, much originality, excellent imagination	
19.	(1) Possesses little or inadequate background information	...	(5) Possesses wide background information and knowledge	
20.	(1) Little curiosity; fails to explore investigate	...	(5) Is keenly curious	
21.	(1) Incorrect view of ability either over or under estimates ability	...	(5) Realizes own capacities and is satisfied; able to evaluate own achievement	
22.	(1) Slow in understanding new ideas and new material	...	(5) Catches on to things quickly; grasps ideas without difficulty	

(1)	(2)	(3)	(4)	(5)
V.	Adjustment to Classroom			
23.	(1) Inattentive; a "dreamer"	...	(5)	Is a good listener; gives rapt attention
24.	(1) Uncooperative; hinders group activities	...	(5)	Cooperates well in organized group activities and functions
25.	(1) Constantly dependent on others for help and directions	...	(5)	Independent; can work independently after initial instructions
26.	(1) Short attention span	...	(5)	Excellent attention span
27.	(1) Restless; can't settle down, needs very varied activities	...	(5)	Able to work quietly for an adequate period of time without becoming restless
28.	(1) Doesn't know what to do next, doesn't get work done; messy; careless	...	(5)	Knows what to do at the right time; completes assigned work; cleans up own space
29.	(1) Displays continual attention getting behavior	...	(5)	Fits into classroom affairs without demanding undue attention from teacher or peers
30.	(1) Rebellious against authority, defiant	...	(5)	Accepts teacher's (authority) role

DEVELOPMENT OF TEACHER EVALUATION SCALES TO PREDICT READING SUCCESS OF PUPILS IN PRIMARY GRADES

This study was a longitudinal one from kindergarten through fourth grade to determine whether measures that are readily available in most school districts, taken in kindergarten by classroom teachers, can predict reading achievement as measured by standardized tests in succeeding grades.

The subjects of the study were 553 children in the kindergarten class of 1964-65 in Ithaca, New York. They were given: the Metropolitan Readiness Test (MRT), including the Goodenough Draw-a-Man Test (DAMT); a ranking by the teacher (TR); and a rating on a composite behavior rating scale (BRS). The measures were used as predictors of achievement on the Metropolitan Achievement Test in first grade and the Stanford Achievement Test in second through fourth grades.

Multiple correlation and regression were used to determine correlations of individual predictors with the criteria, intercorrelations of the predictors, predictions obtained by combining variables, and regression coefficients and constants for computing predicted scores of individuals. Multiple discriminant analysis was used to determine whether high, average, and low achievement

groups were separated by the antecedent variables.

The MRT was by far the best predictor, with coefficients ranging from .63 for first grade to .74 for third grade. BRS and TR had correlations in the .40's and .50's. The DAMT was a poor predictor, with a coefficient of .39 for all four grades. There was considerable overlap among the predictors.

Another correlation analysis used the six MRT subtests instead of the total score, in combination with DAMT, BRS, and TR. The Alphabet and Numbers subtests were the best predictors, with Word Meaning, Matching, and Copying in the same range as Behavior Rating Scale and Teacher Ranking. The Listening subtests was an even poorer predictor than Draw-a-Man Test.

In the analysis using four variables, multiple R's with two to four variables did not increase the predictive power of the MRT, which contributed 40% to 44% of the variance among the grades.

In the nine-variable analysis, two or three predictors on each grade added to the 40% to 46% of the variance which was contributed by the Alphabet subtest. Other variables contributing from 1½% to almost 9% of the variance were: Matching and Teacher Ranking for first grade, Word Meaning and Teacher Ranking for second, Word Meaning and Copying for third grade, and Word Meaning, Copying, and Teacher Ranking for fourth. Multiple R's for nine variables ranged from .681 for first

grade to .772 for third grade.

Multiple discriminant analysis using four predictors revealed only one significant function on each grade level. This function, highly related to all variables on all grade levels, had almost perfect correlation with the MRT for the first two grades.

When the same analysis was applied to nine variables, the first function only was significant for the first two grades. In third and fourth grades, however, the second function also was significant, at the .05 level.

The variables most important to the first function for third grade were Alphabet, Teacher Ranking, and Numbers, and for the second function, Numbers, Matching, and Alphabet.

For fourth grade, Alphabet, Word Meaning, and Listening ranked highest in contribution to the first function. All subtests except Listening, in addition to Teacher Ranking, contributed to the second function.

Because the reliability of short subtests such as those of the MRT is suspect, more importance must be attached to results of the four variable analyses than those of the nine variables.